Predicting Task-specific Performance for Iterative Reconstruction

in Computed Tomography

by

Baiyu Chen

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Ehsan Samei, Supervisor

_____
Joseph Lo

_____
James Colsher

_____
James Bowsher

_____
Rendon Nelson

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in Graduate Program in Medical Physics in the
Graduate School of Duke University

2014

ABSTRACT

Predicting Task-specific Performance for Iterative Reconstruction

in Computed Tomography

by

Baiyu Chen

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Ehsan Samei, Supervisor

_____
Joseph Lo

_____
James Colsher

_____
James Bowsher

_____
Rendon Nelson

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in Graduate Program in Medical Physics in the
Graduate School of Duke University

2014

Abstract

The cross-sectional images of computed tomography (CT) are calculated from a series of projections using reconstruction methods. Recently introduced on clinical CT scanners, iterative reconstruction (IR) method enables potential patient dose reduction with significantly reduced image noise, but is limited by its "waxy" texture and nonlinear nature. To balance the advantages and disadvantages of IR, evaluations are needed with diagnostic accuracy as the endpoint. Moreover, evaluations need to take into consideration the type of the imaging task (detection and quantification), the properties of the task (lesion size, contrast, edge profile, etc.), and other acquisition and reconstruction parameters.

To evaluate detection tasks, the more acceptable method is observer studies, which involve image preparation, graphical user interface setup, manual detection and scoring, and statistical analyses. Because such evaluation can be time consuming, mathematical models have been proposed to efficiently predict observer performance in terms of a detectability index (d'). However, certain assumptions such as system linearity may need to be made, thus limiting the application of the models to potentially nonlinear IR. For evaluating quantification tasks, conventional method can also be time consuming as it usually involves experiments with anthropomorphic phantoms. A mathematical model similar to d' was therefore proposed for the prediction of volume

quantification performance, named the estimability index (e'). However, this prior model was limited in its modeling of the task, modeling of the volume segmentation process, and assumption of system linearity.

To expand prior d' and e' models to the evaluations of IR performance, the first part of this dissertation developed an experimental methodology to characterize image noise and resolution in a manner that was relevant to nonlinear IR. Results showed that this method was efficient and meaningful in characterizing the system performance accounting for the non-linearity of IR at multiple contrast and noise levels. It was also shown that when certain criteria were met, the measurement error could be controlled to be less than 10% to allow challenging measuring conditions with low object contrast and high image noise.

The second part of this dissertation incorporated the noise and resolution characterizations developed in the first part into the d' calculations, and evaluated the performance of IR and conventional filtered back-projection (FBP) for detection tasks. Results showed that compared to FBP, IR required less dose to achieve a threshold performance accuracy level, therefore potentially reducing the required dose. The dose saving potential of IR was not constant, but dependent on the task properties, with subtle tasks (small size and low contrast) enabling more dose saving than conspicuous tasks. Results also showed that at a fixed dose level, IR allowed more subtle tasks to

exceed a threshold performance level, demonstrating the overall superior performance of IR for detection tasks.

The third part of this dissertation evaluated IR performance in volume quantification tasks with conventional experimental method. The volume quantification performance of IR was measured using an anthropomorphic chest phantom and compared to FBP in terms of accuracy and precision. Results showed that across a wide range of dose and slice thickness, IR led to accuracy significantly different from that of FBP, highlighting the importance of calibrating or expanding current segmentation software to incorporate the image characteristics of IR. Results also showed that despite IR's great noise reduction in uniform regions, IR in general had quantification precision similar to that of FBP, possibly due to IR's diminished noise reduction at edges (such as nodule boundaries) and IR's loss of resolution at low dose levels.

The last part of this dissertation mathematically predicted IR performance in volume quantification tasks with an e' model that was extended in three respects, including the task modeling, the segmentation software modeling, and the characterizations of noise and resolution properties. Results showed that the extended e' model correlated with experimental precision across a range of image acquisition protocols, nodule sizes, and segmentation software. In addition, compared to experimental assessments of quantification performance, e' was significantly reduced in

computational time, such that it can be easily employed in clinical studies to verify quantitative compliance and to optimize clinical protocols for CT volumetry.

The research in this dissertation has two important clinical implications. First, because d' values reflect the percent of detection accuracy and e' values reflect the quantification precision, this work provides a framework for evaluating IR with diagnostic accuracy as the endpoint. Second, because the calculations of d' and e' models are much more efficient compared to conventional observer studies, the clinical protocols with IR can be optimized in a timely fashion, and the compliance of clinical performance can be examined routinely.

Contents

List of Tables

List of Figures

Abbreviations

AEC	Automatic exposure control
AIDR	Adaptive iterative dose reduction
ASiR	Adaptive statistical iterative reconstruction
AUC	Area under the receiver operating characteristic curve
CT	Computed tomography
CNR	Contrast-to-noise ratio
ESF	Edge spread function
FBP	Filtered back-projection
GEE	Generalized estimating equation
HU	Hounsfield unit
IR	Iterative reconstruction
IRIS	Iterative reconstruction in image space
LSF	Line spread function
MBIR	Model based iterative reconstruction
MTF	Modulation transfer function
NPS	Noise power spectrum
NPWEi	Non-prewhitening matched filter with eye filter and internal noise
PB	Percentage bias

PE	Percentage error
PRC	Percentage repeatability coefficient
PSF	Point spread function
RC	Repeatability coefficient
RECIST	Response evaluation criteria in solid tumors
ROC	Receiver operating characteristic
ROI	Region of interest
SAFIRE	Sinogram affirmed iterative reconstruction
TTF	Task transfer function
VOI	Volume of interest

Acknowledgements

I would like to express my deepest appreciation and gratitude to my advisor, Dr. Ehsan Samei, who persuaded me into PhD study, and has been extremely supportive and caring since then. Dr. Samei not only introduced me to the field of medical imaging, but also trained me in scientific thinking and writing to help me become an independent researcher. My gratitude also goes to Drs. Joseph Lo and James Colsher for all the advices they have given me in choosing career path. I am also indebted to the rest of my PhD committee, Drs. Rendon Nelson and James Bowsher, for their warm encouragement, insightful comments, and challenging questions. Special thank goes to Dr. James Dobbins, director of my graduate program, for all the guidance and help he has provided me throughout my graduate study.

I greatly appreciate my labmates for the productive collaborations and all the fun we have had in the last five years. Sincere thanks are due to Drs. Xiang Li and Samuel Richard, who unreservedly taught me all the essential skills needed as a graduate student. I would also like to thank Olav Christianson, Justin Solomon, Dr. Jered Wells, Dr. Joshua Wilson, and Dr. James Winslow, for all the stimulating discussions. The experimental parts of this work were only possible with the help of Drs. Juan Carlos Ramirez Giraldo and Troy Zhou. Last but not least, I would like to thank my dear

friends, Yichao Fan, Lijuan Sun, Yakun Zhang, Chu Wang, Lynda Ikejimba, and Yunhui Zhu, for their love and supports throughout my PhD life.

I dedicate this dissertation to my parents, Yuannan Chen and Jihua Chen, who have instilled in me the value of hard work and diligence.

1. Background and Introduction

1.1 Iterative Reconstruction in Computed Tomography

Computed tomography (CT) enables non-invasive examinations of patients with cross-sectional images reconstructed from a series of x-ray projections. The reconstruction is essentially the process of solving an inverse problem: Given the physical measurements of the patient (projections), the reconstruction process converts the measurements back into the information about the patient (cross-sections). To solve the inverse problem, an intuitive approach is to "guess" the patient cross-sections in a feedback mechanism: One starts with an initial guess of the patient cross-sections, and updates the guess in a feedback loop (iteration) by comparing the projections of the guess to the projections of the patient[1]. Multiple iterations are usually required to solve for the patient cross-sections. This feedback mechanism is known as the iterative reconstruction (IR) method.

1.1.1 Theory of IR method

The aforementioned IR scheme is quite simple and straightforward. However, to accurately solve for the complicated human body structure, many details need to be implemented into an IR method. The complexity of the details affects the tradeoff between the computational time and the image quality, and is the choice of the developer of the IR method. Typically, there are five choices that need to be made[2]:

The first choice concerns the parameterization of the object. While the human body is a continuous object, the reconstructed image is a discrete sample of the object. Therefore, an obvious question would be how small the pixels should be. A smaller pixel size allows a better sampling of the patient, but increases the computational burden dramatically. A practical compromise may be to start with a large pixel size that coarsely samples the object, such that a general ideal of the object can be quickly acquired, and then refine the image with smaller pixel size to have a better understanding of the object details[3].

The second choice concerns the modeling of the imaging system. After a guess of the object is made, a forward-projection process converts the guess to projections of the guess. To improve the accuracy of the forward-projection process, many physical properties of the imaging system can be modeled[1], including the scan geometry, the properties of the x-ray source (focal spot size, spatial distribution of beam intensity, and energy spectrum), the interaction between the x-ray and the patient body (scatter and beam-hardening), and the properties of the detector (finite detector size and detector crosstalk). IR methods with these details implemented are known as "model-based IR", which are expected to provide higher image resolution, improved low contrast detectability, and reduced image artifact. However, the computational time for model-based IR is also significantly longer, as the modeling process dramatically increases the complexity of the forward-projection.

The third choice concerns the inclusion of statistical noise models. Due to the impact of quantum noise and electronic noise, the patient measurements (projections) are always distributed around the means. However, the forward-projection process can only predict the mean of the measurements. As a result, even if the guess of the patient cross-section is correct, due to the noise, the projections of the guess can never perfectly match the projections of the patient. To account for this bias, noise models can be included: The quantum noise can be modeled as Poisson noise, and the electronic noise in readout electronics can be modeled as Gaussian noise[4]. After implementing the noise models, a maximum-likelihood (ML) method can be used to look for a guess that maximizes the likelihood of obtaining the true patient projections[5].

The fourth choice concerns the enforcement of desired properties. To get low-noise CT images, a regularization term can be included in IR to penalize image noise[6]. There are various forms of penalties, but they share the same idea that the difference between neighboring pixels is mostly likely due to the noise and should be penalized. With penalties, the reconstructed images can be made smoother and less noisy, but the edges of the body feature may also get penalized and blurred. Therefore, some of the penalties further set a threshold to preserve the edges: When the difference between the neighboring pixels reaches the threshold, the difference is recognized as an edge instead of the noise, and the difference is no longer penalized.

The last choice concerns the algorithm for updating the guess. Between iterations, an algorithm updates the guess and terminates the iterations if a stopping criterion is met. There are multiple algorithms available[7-9] (conjugate gradients, ordered subsets, coordinate descent, etc.). The choice of algorithm affects the convergence rate and the easiness of parallelization.

1.1.2 Examples of currently available clinical IR algorithms

When Sir Godfrey Hounsfield invented CT scanner in 1970s, the first reconstruction method he employed was an IR method. However, at that time, IR was too computational expensive for clinical workflow, and was later replaced on clinical scanners by an analytical method, the filter back-projection (FBP). Recently, the possibility of IR as a clinical reconstruction method has been revisited for three reasons. First, with the developments in modern computing, the computational burden of IR has become less of a problem. Second, compared to FBP, IR has the flexibility of modeling physical conditions that would pose problems for analytical reconstructions, such as irregular gantry orbit, metal implant, and detector crosstalk[1]. Last but most importantly, IR has the ability to significantly reduce patient dose through regularization and statistical model of the quantum noise[10-12], which helps ease the growing public concern about CT radiation dose[13].

All major CT vendors now have IR algorithms available for their clinical scanners. Examples include IRIS[14] (Iterative Reconstruction in Image Space) and

SAFIRE[15] (Sinogram Affirmed Iterative Reconstruction) from Siemens Healthcare, Malvern, PA; ASiR (Adaptive Statistical Iterative Reconstruction) and MBIR[16] (model based iterative reconstruction, marketed as Veo) from GE Healthcare, Waukesha, WI; AIDR 3D[17] (Adaptive Iterative Dose Reduction 3D) from Toshiba Medical Systems Corporation, Otawara, Japan; and iDose[18] from Philips Healthcare, Cleveland, OH.

Because the IR methods described in 1.1.1 with loops of forward-projections are very computationally demanding, all aforementioned clinical IR algorithms have their own approach to reduce the computational time. It is difficult to describe the exact mechanisms behind these algorithms due to their commercial natures. The following describes clinical IR algorithms in three categories and briefly explain each, from the most simplified image-domain method to the most sophisticated model-based method.

The first category is an image-domain method. Unlike traditional IR methods that use forward-projections to iterate between raw data domain and image domain, the image-domain method eliminates the process of forward-projection and iterates within image domain only. To do so, a cross-sectional image is first reconstructed with FBP method. The algorithm then identifies the noise in the image by comparing the fluctuations in the image to a statistical model of the noise, and repeats this noise identification process to iteratively reduce image noise. This method is the most simplified and also the fastest method among all commercial IR methods. IRIS uses this method.

The second category of commercial IR involves both raw data domain and image data domain, but has little interaction between the two. To do so, a loop of noise identification similar to the one performed in the image domain is first performed in the raw data domain. The de-noised raw data is then back-projected into image domain to be further de-noised with a second loop of noise identification process. Compared to the first category, this category pushes noise reduction further. As a compromise, the computational time of this category is slightly longer, but still significantly shorter than traditional IR methods and can be used clinically without interrupting clinical flow. SAFIRE, iDose, ADIR 3D, and ASIR fall in this category.

The aforementioned two categories, while significantly reducing image noise, are still limited in that they continue assuming ideal imaging system like the conventional FBP method. Therefore, IR algorithms in these two categories cannot solve some fundamental problems with FBP, such as beam hardening artifact. The third category, however, includes physical modeling into the algorithm to account for multiple sources of system imperfection, such as the finite focal spot size, the crosstalk between detector elements, the polychromatic x-ray source, and the scattering distribution within human body. This physical modeling is implemented via multiple forward-projections from the image domain to the raw data domain. As a result of the modeling, the noise is further reduced, the resolution is improved, and the beam hardening artifact is reduced. The computational time of this category is the longest among the three, and to some extent

interrupts the clinical flow. However, with the rapid development of modern computing technology, this will not be a concern for long. MBIR falls in this category.

1.1.3 Advantages and disadvantages of clinical IR algorithms

Since the emergence of clinical IR in 2009[11], multiple studies have been conducted to investigate the potential of clinical IR techniques in improving observer performance[10]. Initial studies have reported reduced artifacts[15, 16, 18] (metal, photon starvation, etc.) and a 30-70% reduction in image noise with these IR algorithms as compared to FBP[12, 19-22]. One may assume such attributes to help improve the performance of human observer or any characterization operator that takes the images as input.

However, commercial IR has potential drawbacks too. One drawback is its reported "waxy" noise texture[10, 21, 23] as a result of IR's increased low frequency noise[24, 25]. Because many anatomical features of interest are mainly composed of low frequency components also, IR's waxy texture increases the difficulty of distinguishing anatomical features from noise, and therefore to some extent hampers the perception of the images. Another drawback of IR is its possible loss of image resolution at reduced dose setting due to the tradeoff between image noise and resolution[24], which adds difficulty to the detection or quantification of subtle features. The third drawback of IR is its diversity, as each CT vendor had its own approach to system modeling and computational time reduction. For example, SAFIRE, ASiR, and iDose all involve iterations in both

reconstructed and projection space, each in a different manner, while MBIR involves iterations in the projection space only[12, 26]. As a result, IR is not as standardized as FBP reconstruction: Each IR algorithm needs to be evaluated individually.

1.2 Task-specific Evaluations of IR

Having summarized the general advantages and disadvantages of commercial IR, it is now important to evaluate IR for an optimal tradeoff between the advantages and disadvantages. The evaluations of IR need to be made with diagnostic accuracy as the endpoint. They also need to be performed together with other acquisition parameters, such as dose and slice thickness, to capture possible interactions between IR and other acquisition parameters. Most importantly, the evaluations need to be "task-specific", i.e., with clear definitions of the task type (detection or quantification) and the task properties (nodule size, density, shape).

1.2.1 Evaluating IR for detection tasks

For characterizing detection tasks, the more acceptable method is observer studies. Considering a typical observer study for the detection of lung nodules, images with nodule absent and present are first prepared; graphical user interface is then developed to display one or more images at a time; human observers detect and assign scores to possible nodules according to their confidence for the presence of the nodules; receiver operating characteristic (ROC) curves are generated from the scores; and finally, the areas under the ROC curves (AUC) are calculated to represent the percent accuracy

of the detection performance. Observer study directly reflects the diagnostic accuracy, and therefore is the gold standard of performance evaluation. However, it is time and effort consuming: Even for a single detection task and a single imaging protocol (a certain combination of imaging and reconstruction parameters), a large number of images and observers are required to achieve sufficient statistical significance[27, 28], let alone that IR needs to be evaluated for multiple tasks and along with varying imaging and reconstruction parameters.

To efficiently evaluate detection tasks, mathematical observer models have been developed, which predict human observer performance by modeling the properties of the image, the nodule, and the observer[27-30]. The observer models calculate the detection performance in terms of a detectability index (d'), which can be translated into AUC for further analysis. Previous studies have shown good agreements between the AUC derived from d' and the AUC measured directly with human observers[30-32]. However, certain assumptions, such as system linearity and stationarity may need to be made, which limit the applicability of d' calculations.

CT with IR method is often non-linear in the sense that CT noise and resolution might vary as a function of noise magnitude and task contrast[18, 24]. As a result, to apply observer models for task-specific evaluations of IR, the models must be first extended in their characterizations of image noise and resolution: Instead of assuming constant noise

and resolution, the noise and resolution properties may need to be characterized as a function of noise magnitude and task contrast.

1.2.2 Evaluating IR for quantification tasks

Quantification tasks are tasks that extract numerical information (diameter, volume, density, etc.) from CT images. They are not necessarily evaluated in terms of AUC, but more clinically relevant in terms of the accuracy and precision of the quantification. The accuracy is the degree to which a measured quantity represents its true value. The precision is the degree to which repeated measurements under unchanged conditions show the same outcome, thus indicating the performance in term of discerning changes.

In traditional evaluations of quantification tasks, repeated experimental measurements are usually required. Take the evaluation of lung nodule volume quantification as an example. CT images are first acquired from repeated scanning of an anthropomorphic phantom or a patient; the nodules are then manually or semi-manually segmented from the images with the aid of software; the volumes of the segmented nodules are calculated; and the accuracy and precision of the volume quantification are statistically analyzed. The whole process can take weeks to months[23, 33, 34].

To evaluate IR for quantification tasks more efficiently, mathematical models similar to the detection models can be used. Prior studies have employed models such as

maximum–likelihood estimator[35], classical Wiener estimator[36], and scanning-linear estimator[36] to directly quantify the size, location, and amplitude of the signal. To focus on assessing the performance of the quantification, a recent mathematical model has been developed by Richard and Samei[37] to assess the precision of volume quantification in terms of an estimability index (e'). The e' values correlated well with theoretical quantification precision obtained via a maximum likelihood estimator. However, this prior model was limited in three aspects: 1) for each task (unique nodule size, contrast, shape, etc.), the model needed to be first trained with pre-known precision values, therefore limiting its applicability to a wide range of nodule characteristics; 2) the model assumed ideal quantification that made optimal use of the image for nodule segmentation, while clinical quantification software generally have less efficiency; and 3) the model characterized image noise and resolution properties in a linear fashion, and thus was limited to linear imaging systems.

1.3 Design and Objective of the Dissertation

The objective of this dissertation is to develop a task-specific framework for efficient evaluations of IR performance.

For both detection and quantification tasks, the characterizations of image noise and resolution need to be extended to reflect IR's potential non-linearity. An extended characterization methodology was thus first developed in Chapter 2. The methodology used a previously developed image quality phantom[38] to measure magnitude-

dependent noise properties and contrast-dependent resolution properties in three dimensions. Because such measurements were performed at multiple contrast and noise levels, including high noise and low contrast conditions that were susceptible to errors, a series of procedures were employed to ensure the robustness of the measurements. In addition, simulations were performed to validate the accuracy and precision of this methodology, with guidelines established to ensure <10% measurement error.

Chapter 3 evaluated the performance of IR in detection tasks using the observer models. The extended noise and resolution characterizations developed in Chapter 2 were incorporated into the observer models, such that the detection performance with IR could be mathematically predicted in terms of the detectability index (d'). The d' was calculated for both FBP and IR (IRIS and SAFIRE) images under a range of dose levels, task contrasts, and task sizes. The performance of FBP and IR were then compared in terms of task-specific dose reduction potential and the percentage of tasks exceeding a threshold AUC of 0.9 for a given dose level.

Both Chapters 4 and 5 evaluated the performance of IR for quantification tasks: Chapter 4 evaluated IR performance using experimental methods, while Chapter 5 evaluated IR performance using mathematical models. To evaluate IR experimentally, an anthropomorphic chest phantom with synthetic nodules of two sizes were scanned using a range of protocols, including FBP and two IR algorithms (ASiR and MBIR). The nodule volumes were quantified with two segmentation software. The accuracy and

precision of the volume quantification were calculated for each protocol, nodule size, and segmentation software. The impact of IR on the accuracy and precision of quantification was then evaluated with respect to nodule size, dose level, reconstruction slice thickness, and software.

Chapter 5 mathematically predicted the quantitative performance of IR via the estimability index (e'). The e' model employed in this dissertation work was an extension of the prior model[37]: The nodule-specific training process was replaced with a more efficient, physically-based modeling process; the imperfection of the quantification software was modeled in terms of the discrepancy between the actual nodule and the segmentation software's expectation; and the characterizations of noise and resolution was extended to reflect IR's non-linearity. The new e' model was then used to predict the precision values experimentally assessed in Chapter 4, such that e' was validated against experimental results across a range of acquisition protocols, nodule sizes, and segmentation software.

2. Generalized Measurements of CT Noise and Resolution*

2.1 Introduction

The noise and resolution properties of computed tomography (CT) images are of great importance to diagnostic performance. Quantifying these properties has many applications, such as benchmarking image quality across systems, optimizing acquisition and reconstruction parameters to improve dose efficiency, and predicting observers' performance. Conventionally, these properties are quantified with methods adopted from planar x-ray imaging, where the noise is measured on structure-free images in terms of noise power spectrum (NPS)[39, 40], and the spatial resolution is measured on images of high contrast slits[41], edges[42], or bars[43] in terms of modulation transfer function (MTF). In their conventional implementations, both the NPS and the MTF methodologies assume a linear system and are, therefore, independent of object contrast and noise magnitude.

With nonlinear iterative reconstruction (IR), however, the system noise and resolution may vary with noise magnitude and object contrast[18, 24, 25]. As a result, we have extended the conventional concepts of NPS and MTF to magnitude-dependent NPS and contrast-dependent task transfer function (TTF, also referred to as "task-based MTF" in previous studies), which characterize the noise and resolution properties at a specific

* This chapter is based on a manuscript submitted to *Medical Physics*.

noise magnitude and object contrast level[24, 25, 44]. Both magnitude-dependent NPS and contrast-dependent TTF have been shown to be of experimental value[45] and effective for predicting human observer performance[44, 46]. However, their experimental measurements are challenging for several reasons. First, to be relevant to the potential non-linearity of IR, the measurements need to be performed at multiple contrast and noise levels. Second, performing these measurements in a clinical setting must be able to be done efficiently. Third, measurements made under high noise and low contrast conditions are susceptible to errors, so the effects of noise should be mitigated. Finally, the accuracy and precision of the measurement methodology need to be fully characterized to establish guidelines for robust measurements.

The conventional characterizations of noise and resolution are also limited in their dimensionality. Conventional measurements are often performed in the axial plane (bi-dimensional), which may be sufficient when a CT dataset is viewed in a slice-by-slice fashion, but insufficient for volumetric operations such as three-dimensional surface rendering, multi-planar reformation, and volume quantification. For the increasing applications of volumetric CT, it is important to measure the noise and resolution properties for all spatial dimensions.

To address the aforementioned limitations, this study aimed to develop a CT noise and resolution measurement technique that is relevant for non-linear CT systems and volumetric (3D) imaging. This objective was achieved in two steps: 1) a phantom-

based measurement technique was developed to measure noise and resolution in an efficient and 3D manner; and 2) the robustness of the technique at challenging imaging condition was validated through simulations, with guidelines established for achieving targeted accuracy and precision. In addition, the methodology was applied to commercial CT systems to ascertain its clinical applicability.

2.2 Methods and Materials

2.2.1 Theory

The NPS describes the square of the image noise (variance) in the Fourier domain as a function of spatial frequency. For a linear system, the functional shape of NPS is independent of noise magnitude. In this study, to capture the possible dependency of the shape of the NPS as a function of noise magnitude with non-linear algorithms, the NPS is defined to depend on the magnitude of the image noise. The NPS is also defined in three dimensions for volumetric applications, as

$$NPS_{3D}(u,v,w;N) = \frac{1}{V}\left|\mathcal{F}_{3D}\{I(x,y,z;N) - \overline{I(x,y,z;N)}\}\right|^2, \quad (2.1)$$

where I(x,y,z;N) is the intensity of the volume at the noise magnitude N; u, v, and w are spatial frequencies; and V is the volume of the VOI. The area under the NPS curve reflects the magnitude of the noise, and the shape of NPS reflects the texture of the noise in terms of the spatial correlation of pixel values.

The second figure of merit, TTF, is an extension of the linear system concept of the modulation transfer function (MTF) to accommodate both linear and non-linear

systems. While MTF describes the resolution of a linear system that is independent of the imaging task, TTF describes the resolution of the system for a given "task," i.e., for a specific object contrast and a specific background noise level. The idea is that the resolution of non-linear systems depends on the task, and therefore cannot be sufficiently characterized with one MTF curve. Instead, multiple TTF curves are employed to characterize the system resolution in multiple sub-spaces of tasks, each containing tasks of similar contrast and noise levels. The resolutions of the tasks within the same sub-space are sufficiently close to be considered constant, such that the sub-space is quasi-linear. This quasi-linear approximation is consistent with the treatment of non-linear resolution conditions in prior works[45, 47, 48]. For linear systems, the resolution of the entire task space can simply be characterized with one single TTF, which is identical to the MTF.

The TTF is calculated as the Fourier transform of the line spread function (LSF) with respect to the specific object contrast and image noise. The LSF can be derived from the edge spread function (ESF), the system's response to an edge impulse[49]. Because a single TTF only represents the system resolution at a specific task contrast and noise level, a library of TTFs may be used to characterize the resolution for a relevant range of non-linear operating space. A single TTF also only represents the resolution along the direction perpendicular to the edge. To measure a 3D TTF that fully samples the system's spatial resolution, multiple TTFs need to be measured along various directions.

2.2.2 Measurement technique

The 3D NPS and in-plane TTF are measured from a dedicated phantom developed at our institution, so-called Mercury Phantom[38] (Figure 2.1). The phantom is composed of four cylindrical sections of various diameters (16, 23, 30, and 37 cm) and three tapered connecting sections, all made of polyethylene (-90 HU at 120 kVp). Each cylindrical section is further divided into a 3-cm long uniform subsection for NPS measurements and a 3-cm long subsection with 2.54-cm-diameter rod inserts for in-plane TTF measurements. The inserts are consist of four contrast levels (air, -910 HU; Teflon, 1000 HU; Acrylic, 205 HU; and polystyrene, 50 HU; all at 120 kVp). The TTF along the axial direction is measured with a supplemental phantom, as shown in Figure 2.2. The phantom, denoted as Z-res Phantom hereinafter, is composed of three 10-cm cylindrical sections of different materials (polyethylene, Acrylic, and Teflon). The plane edges at the interfaces of sections enables the measurements of axial TTF at four contrast levels (air to polyethylene, 910 HU; polyethylene to Acrylic, 205 HU; Acrylic to Teflon, 795 HU, and Teflon to air, -1910 HU; all at 120 kVp). The phantom is 5°-tilted in the sagittal plane to yield pre-sampled TTF for slices up to 8-mm thick. The compact arrangement of the rod inserts in Mercury Phantom and the edge planes in Z-res Phantom allows TTF to be measured at four contrast levels in a single scan, which greatly improves the efficiency of the measurements.

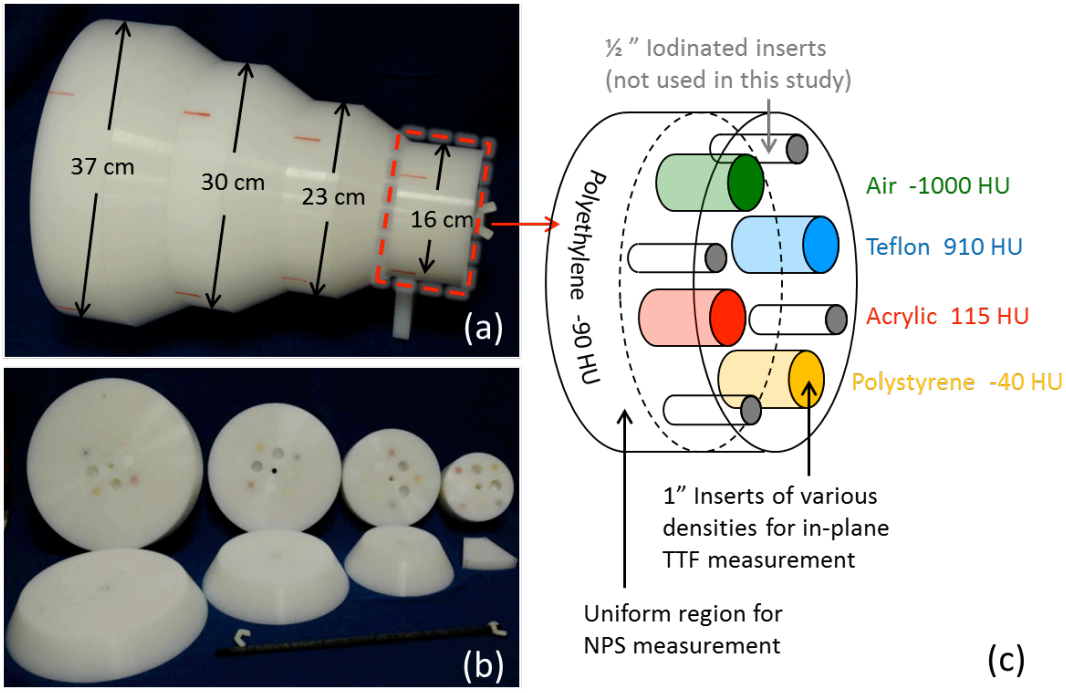

Figure 2.1: (a,b) Mercury Phantom is consists of four cylindrical sections of increasing diameters and three tapered connecting sections. (c) The four cylindrical sections can each be divided into two subsections: one uniform subsection for NPS measurements, and one subsection with rod inserts for in plane TTF measurements.

Figure 2.2: A supplemental phantom for TTF measurements along the axial direction. The phantom consists of three cylindrical sections of different materials, the interfaces of which provide plane edges for TTF measurements.

To measure NPS, multiple regions of interest (ROI) are extracted from slices across a uniform section of the Mercury Phantom (Figure 2.3 (a)). The ROIs are arranged concentrically to address possible non-stationarity of noise across the field of view. The distance of the ROIs from the phantom center is flexible, but preferably at the same depth as the rod inserts shown in Figure 2.1 (c), such that NPS and TTF are measured at the same depth. The physical size of the ROI is based on the size of the foveal vision, which is the central 2-3 degrees of the view that has the best visual acuity[50]. Large scale variations within each ROI are removed with a second order polynomial fit prior to the NPS calculation, as such non-uniformities are generally not considered noise. ROIs of

successive slices are then combined to yield volumes of interest (VOI) (Figure 2.3 (b)). 3D NPS are calculated from each VOI according to Eq. 2.1, and averaged over multiple locations.

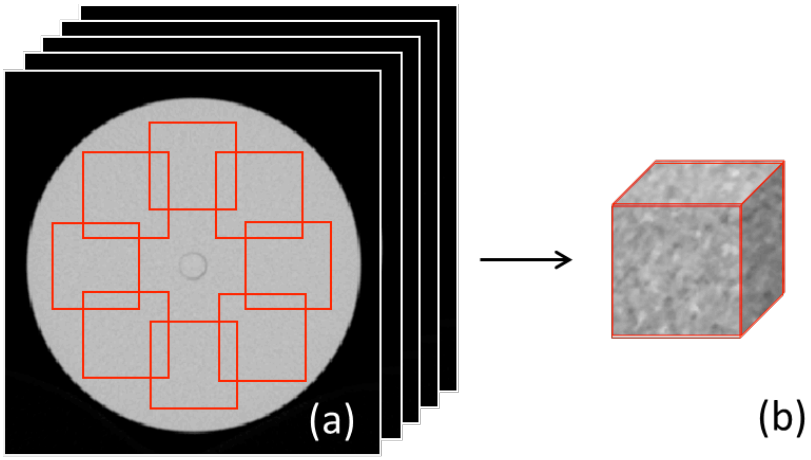

Figure 2.3: (a) ROIs are concentrically extracted from the uniform region of the phantom image. (b) ROIs from successive slices are combined to yield VOIs for 3D NPS measurements.

The 3D TTF is calculated from two components: The TTF that represents the in-plane resolution, TTF_{xoy}, and the TTF that represents the axial resolution, TTF_z. A complete flowchart of the 3D TTF measurement is shown in Figure 2.4. TTF_{xoy} is calculated from the Mercury Phantom, using the circular edges in the cross-sectioned images of the rod inserts. An ROI twice the size of the insert diameter is extracted around the center of the insert, providing balanced space on both sides of the circular edge to capture the full extent of the ESF. Assuming isotropic in-plane resolution, the Hounsfield Units (HU) of all pixels in the ROI are plotted as a function of their distance from the center of the insert to form the "raw" ESF, as shown in Figure 2.4 (e). The raw

ESF, however, are generally too noisy to be used directly for the TTF calculation, especially those measured from low contrast edges with low exposure (i.e., relatively high image noise). Therefore, four de-noising steps are applied to reduce the noise in the ESF:

1) ROIs of an insert are averaged over consecutive slices to reduce the image noise prior to calculating the raw ESF. In doing so, the edges of the insert in all slices need to be precisely aligned to avoid edge blurring due to averaging. This is largely ensured by a precise alignment of the Mercury Phantom with the calibrated laser indicators prior to the scans.

2) The raw ESF spatial binning is super-sampled and rebinned into one tenth of the pixel size to ensure a good uniform sampling without loss of resolution[49].

3) The rebinned ESF is further smoothed with an edge-preserving smoothing algorithm adapted from a previous work[51]. The algorithm iteratively smoothes the ESF curve y_i by finding a new set of points Y_i that minimizes

$$\chi^2 = \sum_i (y_i - Y_i)^2 , \tag{2.2}$$

while satisfying the constraint that the new ESF Y_i is monotonic. In our study, to help preserve the edge enhancement feature seen in some iteratively reconstructed images, the monotonic constraint was extended in the region of edge enhancement to tolerate minor fluctuations between consecutive points. The fluctuation varied between 0-3% of the edge magnitude, depending on the extent of the edge enhancement.

4) After differentiating the ESF to obtain the LSF, the LSF is smoothed with a Hann function with the same peak location as that of the LSF. This helps to remove the noise on the tails of the LSF while preserving peak sharpness. The TTF_{xoy} is calculated from the de-noised LSF.

The TTF_z is calculated from the Z-res Phantom. The interfaces of three sections provide plane edges for TTF_z measurements. VOIs are segmented from the image dataset to enclose the plane edges. Virtual planes are fitted to the edges. The raw ESF is plotted as the voxel intensity against the signed distance between the voxel and the virtual plane. TTF_z is then calculated from the raw ESF following the same series of de-noising procedures as in the calculation of TTF_{xoy}.

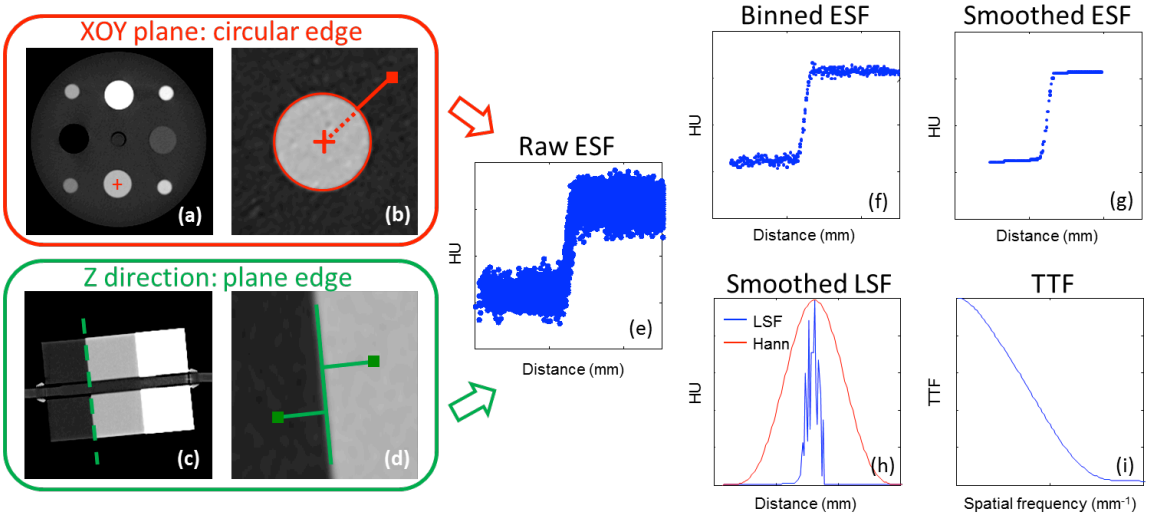

Figure 2.4: 3D TTF is acquired separately in terms of TTFxoy and TTFz. (a,b) To acquire the raw ESF for TTFxoy, the center of the rod insert is located from the averaged slices, with pixels around the center plotted for their HU against their distance from the center to yield (e) the raw ESF. (c,d) Along axial direction, the slanted edge of the Z-res Phantom is fitted with a virtual plane, with voxels on both sides of the plane plotted for

their distance from the virtual plane against their HU to yield (e) the raw ESF. (f-i) To acquire the TTF from the raw ESF, the raw ESF is (f) re-binned, (g) smoothed with an edge preserving algorithm, (h) differentiated into the LSF, smoothed with a Hann window, and (i) Fourier transformed to obtain the TTF.

2.2.3 Validation

The NPS measurement was validated via simulated ROIs with a known noise spectrum. The pixel size was 0.82x0.82 mm^2, assuming a 420-mm reconstructed field of view for the Mercury Phantom. The size of the ROIs was 64x64, corresponding to the foveal field of view at 100 cm viewing distance. Each ROI contained randomly simulated Gaussian white noise with a standard deviation of 50 HU. The true NPS, NPS_{true}, was thus a flat spectrum with energy equal to 50 HU and a cutoff frequency at the Nyquist frequency. The NPS measured with our technique, $NPS_{measured}$, was the average of the NPS measured from multiple ROIs. To validate our method, $NPS_{measured}$ was compared to NPS_{true} in terms of percent error (PE_{NPS}), as

$$PE_{NPS} = \frac{\iint \left| NPS_{measured}(u,v) - NPS_{true}(u,v) \right| dudv}{\iint NPS_{true}(u,v) dudv} = \frac{\iint \left| NPS_{measured}(u,v) - NPS_{true}(u,v) \right| dudv}{\sigma^2}. \qquad (2.3)$$

To assess the accuracy and precision of the NPS measurement, the calculation of PE_{NPS} was repeated multiple times, with the accuracy calculated as the mean PE_{NPS} over repeats, and the precision as the standard deviation of the PE_{NPS} over repeats. The number of the repeats was chosen to be 300, which provided a good balance between computational time and representing the population. In addition, because PE_{NPS} might

be affected by the number of the ROIs employed in the $NPS_{measured}$ calculation, the accuracy and precision of the NPS measurement were further calculated as a function of the number of the ROIs (8, 40, 80, 160, 240, and 320) to derive the minimum number of ROIs required to achieve <10% error.

The TTF measurement was also validated via simulated ROIs. An ROI with a uniform disk in the center was first simulated to represent the cross-section of the rod insert. The pixel size of the ROI was the same as the pixel size of Mercury Phantom images reconstructed with a 420-mm field of view. The diameter of the disk was the same as the rod inserts in the Mercury Phantom. The width of the ROI was twice the diameter of the disk. The edge of the disk was blurred according to the partial volume effect. No noise was simulated in the ROI, so that the TTF measured from the noise-free ROI represented the true TTF, TTF_{true}. Next, simulated noise background was superimposed onto the noise-free ROI. Three types of noise texture were simulated to represent three CT reconstruction algorithms, including FBP and two IR algorithms (Iterative Reconstruction in Image Space, IRIS, and Sinogram Affirmed Iterative Reconstruction at strength 5, SAFIRE5; Siemens Healthcare, Malvern, PA). The TTF measured from the noisy ROI, $TTF_{measured}$, was compared to TTF_{true} in terms of the percent error (PE_{TTF}), as

$$PE_{TTF} = \frac{\iint \left| TTF_{measured}(u,v) - TTF_{true}(u,v) \right| dudv}{\iint TTF_{true}(u,v) dudv}. \quad (2.4)$$

The accuracy and precision of the TTF measurement were calculated from 300 repeated PE_{TTF} measurements. To assess the dependency of the accuracy and precision on the contrast to noise ratio (CNR) of the disk (rod insert), the aforementioned processes were performed at eight disk contrast levels with a fixed noise (noise = 25 HU; disk contrast = 25, 75, 125, 175, 225, 275, 325, and 375 HU; CNR = 1, 3, 5, 7, 9, 11, 13, and 15) to empirically determine the minimum CNR required to achieve an error less than 10%.

For a low contrast insert with high image noise, the minimum CNR derived above may not be achievable in a single slice. In that case, multiple slices across the insert can be averaged to lower the noise and thus to increase the CNR. The number of slices needed to achieve the minimum CNR was also derived as

$$\text{number of slices} = \left(\frac{CNR_{\min}}{Contrast\ /\ Noise} \right)^2 . \tag{2.5}$$

2.2.4 Experimental measurements

As a demonstration, the methodology was used to measure the noise and resolution properties of a commercial CT system. Both the Mercury and the Z-res phantoms were scanned on a multi-detector CT scanner (SOMATOM Definition Flash CT scanner, Siemens Healthcare, Malvern, PA) with 64x0.6-mm collimation, 0.6 pitch, 1-s rotation time, 500-mm scan field of view, and five dose levels (64.8, 27.0, 13.5, 6.7, and 3.4 mGy). The images were then reconstructed with 420-mm reconstruction field of view, two slice thicknesses (0.6 and 1.5 mm), and three reconstruction algorithms (FBP,

IRIS, and SAFIRE5). Both IRIS and SAFIRE5 are commercially available iterative reconstruction algorithms. IRIS iterates in the image domain to reduce image noise while maintaining spatial resolution[11, 12, 14]. SAFIRE5 iterates in both raw data and image domain to reduce image noise and image artifacts[11, 12, 15]. The NPS and TTF of all three reconstruction algorithms were measured from the phantom images using the validated methodology.

2.3 Results

2.3.1 Validation of NPS measurements

The mean and standard deviation (depicted as error bars, extremely small in this case) of PE_{NPS} over 300 repeats are shown in Figure 2.5, which represent the accuracy and precision of the NPS measurements, respectively. The accuracy improves considerably as the number of the ROI increases. With standard deviation depicted in Figure 2.5 being extremely small, the precision of the NPS measurements is extremely good. The results indicate that to achieve an NPS measurement with a percent error of <10%, a minimum number of 65 ROIs would be required.

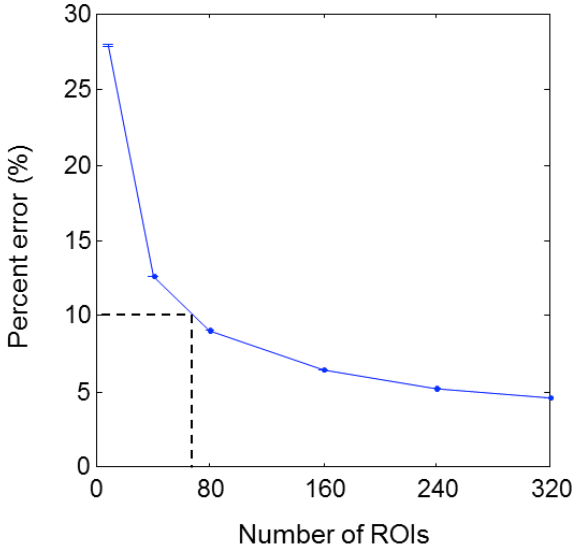

Figure 2.5: The percent error of the NPS measurements averaged over 300 repeated simulations.

2.3.2 Validation of TTF measurements

The mean and standard deviation (error bar) of PE_{TTF} over 300 repeats are shown in Figure 2.6, which represents the accuracy and precision of the TTF measurements, respectively. The accuracy and precision improves considerably as the disk CNR increases. The results indicate that the performance of all three reconstruction algorithms converge and plateau at CNR=15, which provides reasonable accuracy and precision (10±3% error).

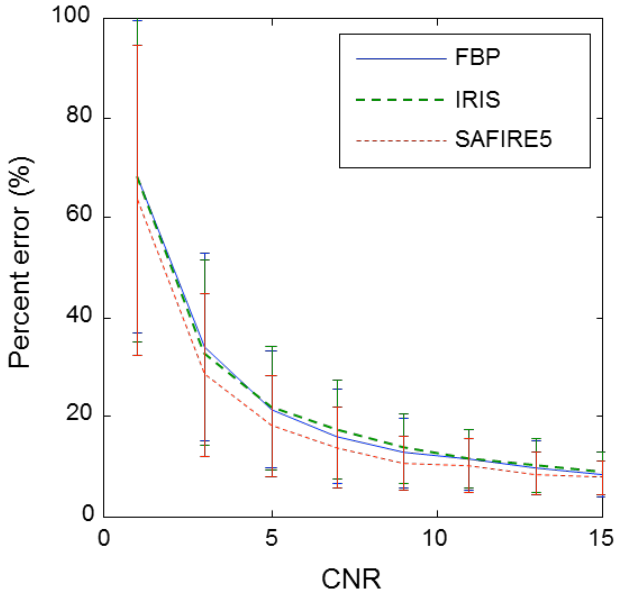

Figure 2.6: The averaged percent error of the TTF measurements as a function of the CNR of the disk.

In case the CNR of 15 is not achievable in a single slice, as noted earlier, multiple slices across the insert can be averaged to increase the CNR. The number of slices needed to achieve a CNR of 15 is plotted in Figure 2.7 for varying contrast and noise levels. With 0.625-mm slice thickness, the 30-mm long rod insert in the Mercury Phantom can accommodate ~40 slices for slice averaging. Therefore, a reliable TTF measurement can be achieved for all combinations of contrast and noise levels above the iso-slice line of 40 in Figure 2.7.

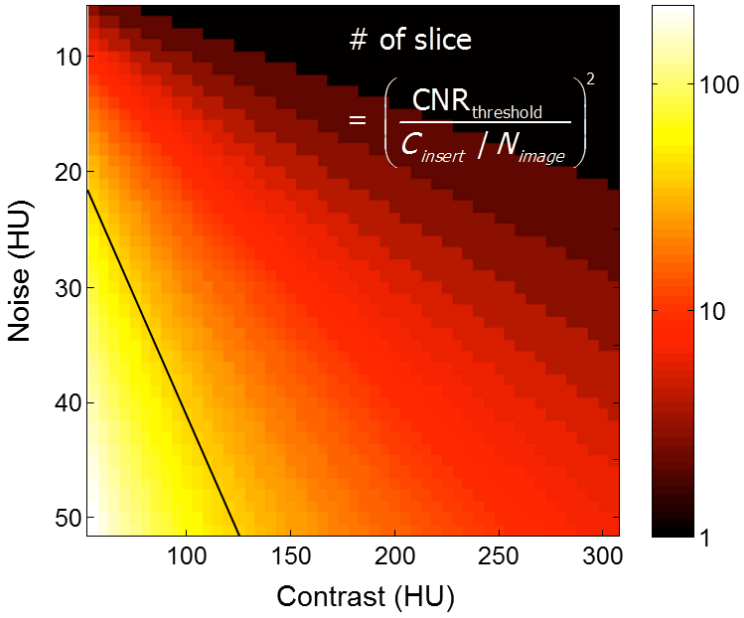

Figure 2.7: The number of slices required to achieve a CNR of 15 by slice averaging. The iso-slice line corresponds to 40 slices.

2.3.3 Experimental measurements

Examples of measured 3D NPS and TTF are shown in Figure 2.8. To better visualize and compare the NPS and TTF measured at various parameters, Figure 2.9 and Figure 2.10 show the radially averaged in-plane figures of merit. Figure 2.9 (a) shows the ROIs from the 23-cm section of the Mercury Phantom. Compared to FBP, the ROIs of IR showed noise of reduced magnitude (IRIS, 30% and SAFIRE5, 55%), corresponding to the lower NPS magnitudes in Figure 2.9 (b). The NPS also demonstrate IR's "waxier" texture with reduced NPS peak frequencies (FBP, 0.25 mm^{-1}; IRIS, 0.23 mm^{-1}; and SAFIRE5, 0.16 mm^{-1}).

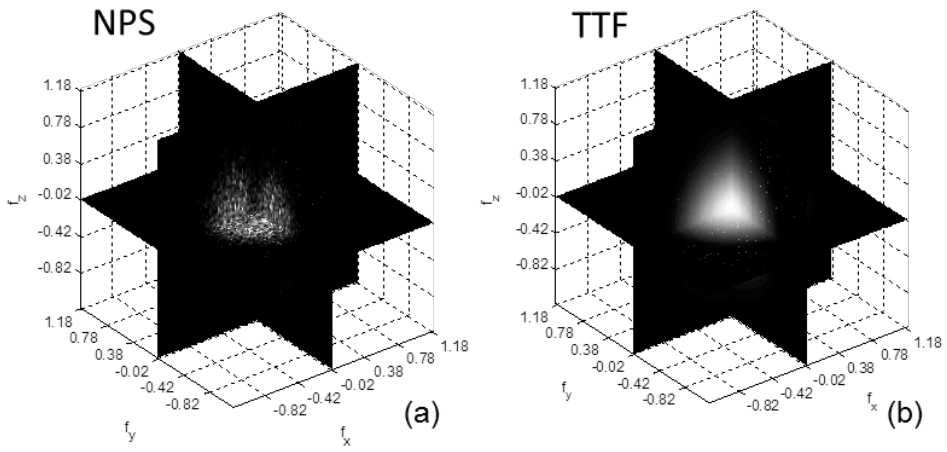

Figure 2.8: Examples of (a) 3D NPS and (b) 3D TTF.

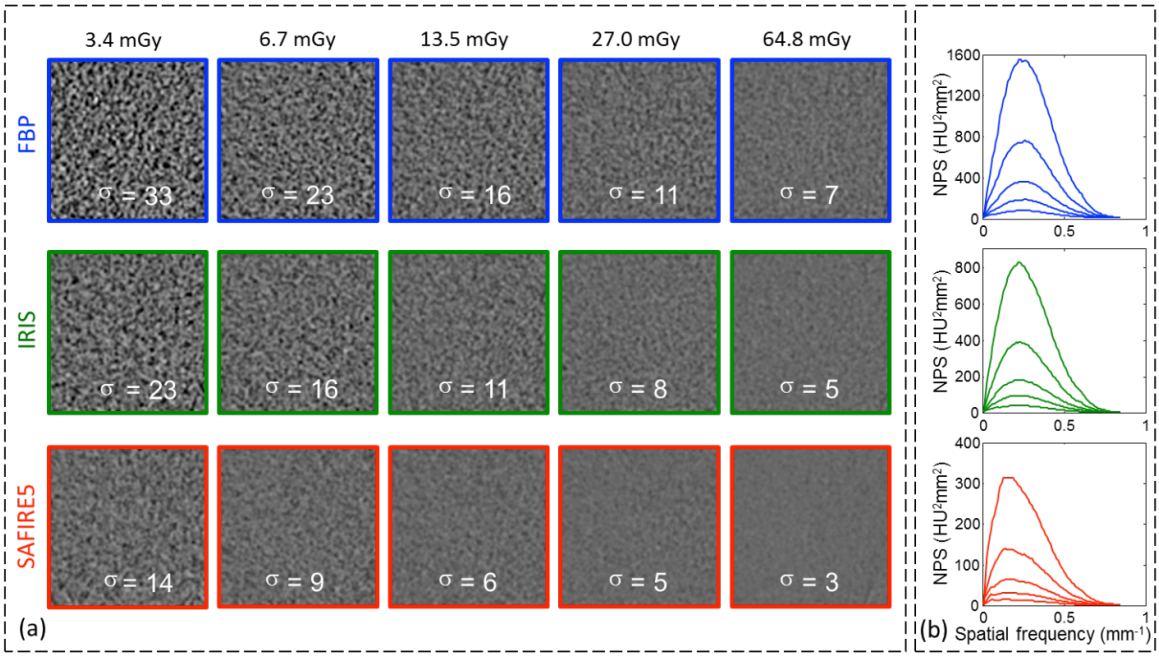

Figure 2.9: (a) ROIs extracted from the 23-cm section of the Mercury Phantom, imaged at five dose levels and reconstructed with three algorithms. (b) Radially averaged in-plane NPS corresponding to each ROI.

Figure 2.10 (a) shows the ROI of the inserts in the 23-cm section of Mercury Phantom, imaged at a dose level of 64.8 mGy. Edge enhancements were visually observed around the rod inserts in all ROIs. Improvement in edge definition with IR was observed in ROIs of low contrast inserts. Figure 2.10 (b) shows the corresponding TTF_{xoy} curves, which reflect the edge enhancements observed in Figure 2.10 (a) with "bumps" near DC frequency, and reflect IR's improvement in edge definition with higher 50% TTF frequency, f_{50} (FBP: polyethylene, 0.41 mm^{-1}, other inserts, 0.36 mm^{-1}; IRIS: polyethylene, 0.42 mm^{-1}; other inserts, 0.37 mm^{-1}; SAFIRE5: polyethylene, 0.46 mm^{-1}, Acrylic, 0.4 mm^{-1}, other inserts, 0.37 mm^{-1}). Note that for SAFIRE5, the f_{50} for polyethylene insert is much higher than the f_{50} for other inserts, reflecting SAFIRE5's non-linearity. The magnitude of this contrast dependency was also confirmed via comparing the areas under the TTF curves: The area under the polyethylene curve was ~30% larger than the area under other curves, much greater than the magnitude of the error associated with the TTF calculation (10±3% according to Figure 2.6).

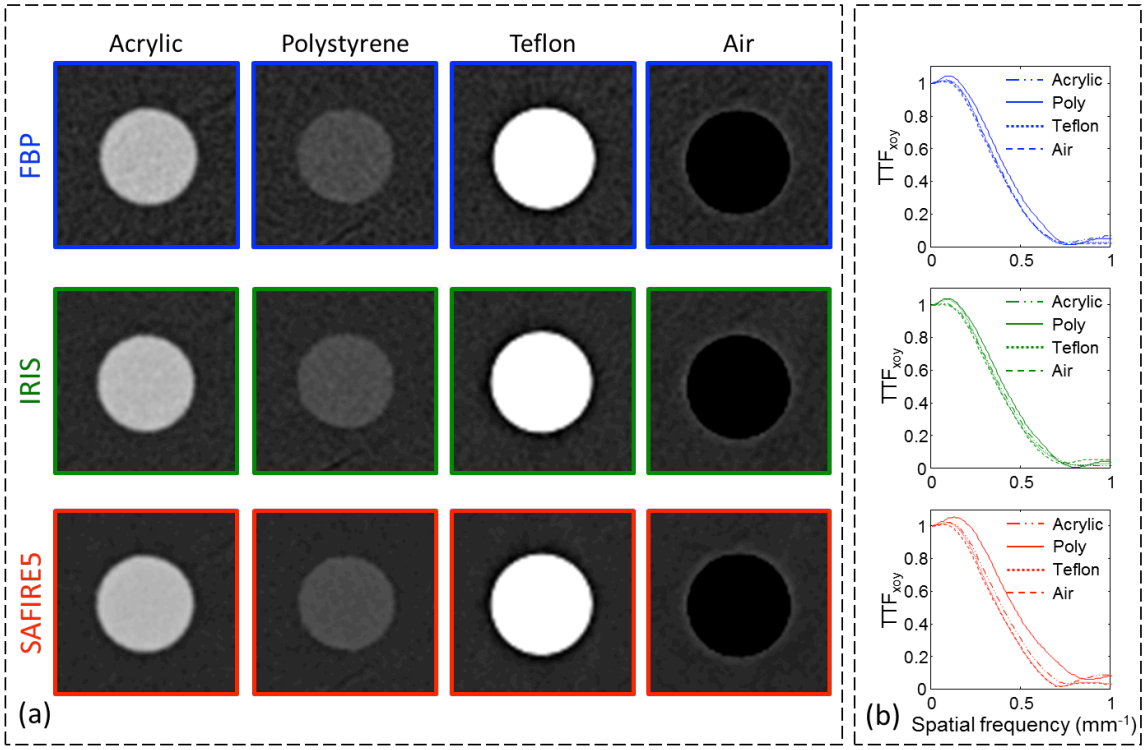

Figure 2.10: (a) ROIs extracted around the inserts in the 23-cm section of the Mercury Phantom, imaged at 64.8 mGy and reconstructed with three algorithms. (b) Radially averaged TTFxoy corresponding to each ROI.

Figure 2.11 shows the edges of the Z-res Phantom (magnified to show the stepwise appearance of the edge through successive slices, full view provided in Figure 2.4) and corresponding TTF_z curves at both 0.6 and 1.5 mm slice thicknesses. The ROIs at thinner slice thickness showed sharper edge transitions, thus better axial resolution. The TTF_z captured the better resolution at thinner slice thickness by showing higher overall values (f_{50} at 0.6 mm slice thickness, 0.35-0.79 mm^{-1}; f_{50} at 1.5 mm slice thickness, 0.22-0.3 mm^{-1}). In addition, the TTF_z captured the improvements in axial resolution as a result of

the flying focal spot technique[52] employed in the scanning: For both slice thicknesses, the TTF_z extended to around twice the Nyquist frequency (0.6 mm slice thickness, $f_{Nyquist}$ = 0.83 mm^{-1}; 1.5 mm slice thickness, $f_{Nyquist}$ = 0.33 mm^{-1}). The impact of contrast on axial resolution depends on the reconstruction algorithm. For FBP and IRIS, the three lower contrast levels showed similar TTF_z (f_{50} at 0.6 mm slice thickness: air-polyethylene, 0.49 mm^{-1}; polyethylene-acrylic, 0.52 mm^{-1}; and acrylic-Teflon, 0.48 mm^{-1}); only the highest contrast level at Teflon-air interface showed a deteriorated TTF_z (f_{50} at 0.6 mm slice thickness: 0.35 mm^{-1}), possibly due to the increased partial volume effect at high contrast levels. For SAFIRE5, all contrast levels showed different TTF_z (f_{50} at 0.6 mm slice thickness: air-polyethylene, 0.64 mm^{-1}; polyethylene-acrylic, 0.79 mm^{-1}; acrylic-Teflon, 0.49 mm^{-1}; and Teflon-air, 0.42 mm^{-1}), with higher contrast levels showing lower TTF_z, i.e., worse resolution. Similar to TTF_{xoy}, the magnitudes of these contrast dependencies were verified by comparing the areas under the TTF_z curves. This strong dependency of resolution on the contrast level reflected SAFIRE5's non-linearity, and substantiated the importance of evaluating resolution with respect to task properties.

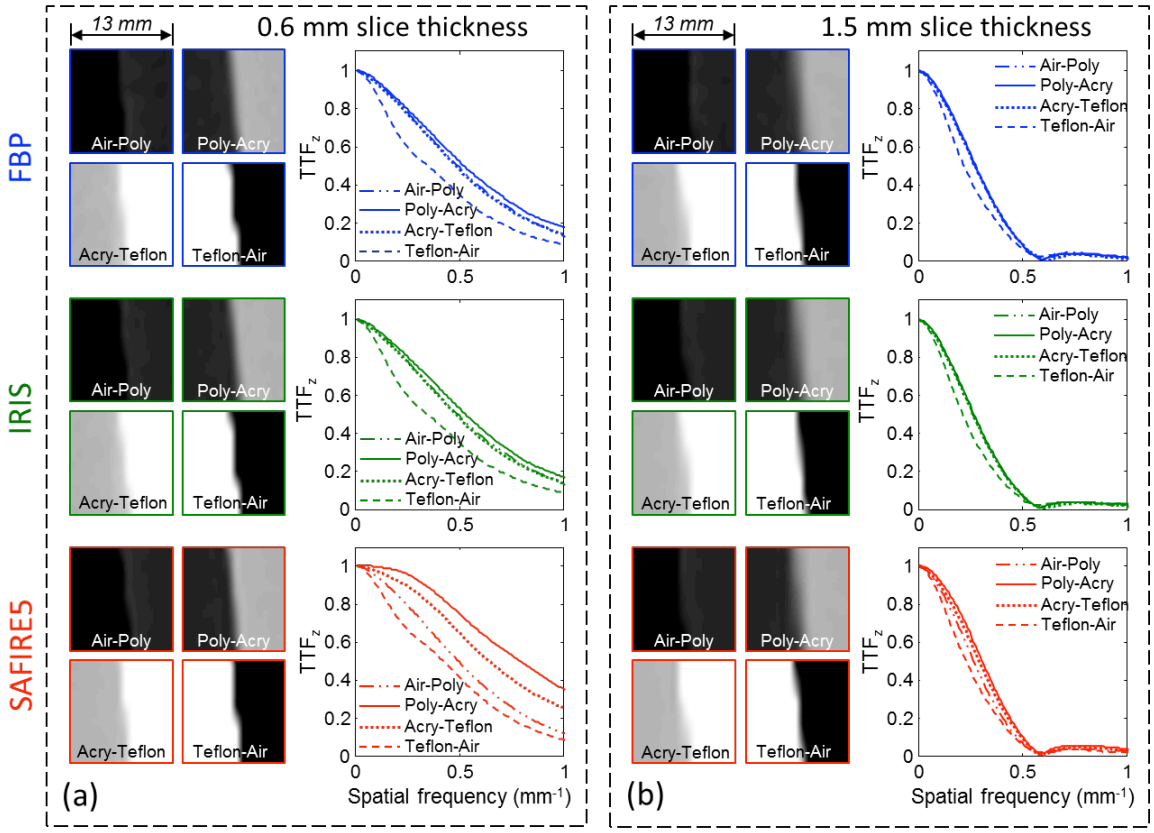

Figure 2.11: The flat-plane edges of Z-res Phantom (magnified sagittal view) and corresponding TTFz curves at (a) 0.6 mm and (b) 1.5 mm slice thicknesses.

2.4 Discussion

Characterization of CT noise and resolution properties plays a central role in image quality assessment, imaging parameter optimization, and performance prediction. For non-linear iterative reconstructions, the conventional characterization of noise and resolution properties need to be extended to varying noise and contrast levels, including challenging conditions of clinically-relevant low contrast and high noise levels. To address this challenge, this work introduced a phantom-based methodology to ensure robust measurements at challenging conditions, with established guidelines to achieve target accuracy and precision levels. In addition, the characterization of noise and resolution was extended to three dimensions to accommodate volumetric operations such as three-dimensional surface rendering and CT volume quantification.

The validation of the NPS measurements showed that to achieve a percent error of less than 10%, a minimum of 65 64x64 ROIs is required. Although this requirement was deduced based on a noise of 50 HU in magnitude, the percent error would be applicable to other noise magnitudes because its definition is already normalized by the noise magnitude (Eq. 2.3). For the Mercury Phantom, which has a 30-mm long uniform subsection for NPS measurements, the requirement can be achieved by satisfying the relationship of

$$\frac{30}{d} * n \geq 65, \tag{2.6}$$

where n is the number of ROIs extracted from each slice and d is the slice thickness of the images. In case the number of slices in the uniform subsection is limited by the slice thickness, more ROIs can be extracted from each slice to meet the criterion by allowing overlap.

Our 3D TTF is measured under two assumptions, the isotropy of in-plane resolution, and the independence between in-plane resolution and axial resolution. One way to avoid these assumptions is to employ spherical phantoms that directly measure resolution along arbitrary directions with edges within a small solid angle[53-55]. The spherical phantom method, however, requires relatively large spheres to ensure sufficient sampling density along each direction. This can be difficult when there is a need to arrange multiple spheres of various attenuations together for TTF measurements. Another alternative is to use point inserts to measure 3D resolution properties from point spread function (PSF)[55, 56]. The point insert method is limited in that it requires inserts with sufficiently high density to generate enough signal magnitude for PSF and is therefore not applicable to TTF measurements at low contrast levels. Compared to sphere and point methods, our approach provides a trade-off between task-based specificity and phantom construction feasibility.

A similar circular-edge-based technique for in-plane TTF measurements was employed in a prior study[24]. Compared to the prior study, our method employed further noise reduction procedures to ensure robust TTF measurements for high noise and low

edge contrast conditions. More importantly, the prior study validated the TTF measurements at high CNR level only. Our study validated the TTF measurements at a range of CNR levels, especially at low CNR levels where the TTF measurements are challenged. Finally, the prior study only validated the accuracy of TTF measurements, while our study employed repeated simulations to validate both accuracy and precision, with guidelines provided for compliance.

Most clinical tasks are represented by a continuous range of contrast values. However, due to practical limitations, only four discrete contrast levels were employed in this study for the calculations of TTF_{xoy} and TTF_z. For low-contrast tasks most relevant to nodule detection, the resolution of the tasks varies slowly such that our TTF measured with Acrylic and polystyrene inserts may be representative of tasks at similar contrast levels. For high contrast tasks such as bone and contrast angiography studies, there is a higher degree of approximation, but provided that the contrast level for the TTF measurement matches the contrast of the task, the TTF may be considered as representative of the overall resolution of the task.

While our characterizations of noise and resolution were performed in Fourier domain, it is worthwhile mentioning that the characterization can also be performed in spatial domain in terms of covariance matrices. The main strength of the spatial domain approach is that it is not limited to linear systems, and can therefore be applied directly to iterative reconstruction[57]. However, the spatial domain approach is practically

challenging as it requires hundreds of repeated scans to calculate the covariance matrices for a single task. Our Fourier approach, on the other hand, requires only a few scans of the Mercury Phantom to characterize the task space, and can therefore be implemented into routine quality assurance without interrupting the clinical workflow. Furthermore, a previous study has shown that our Fourier approach provide a reasonable model of human observer performance with non-linear iterative reconstruction across a wide range of systems and dose levels[44, 46].

Notwithstanding the findings of the study, certain qualifications need to be acknowledged. First, the validation of the NPS was performed based on white noise, while CT noise is correlated. This was done to provide a universal validation independent of the spectrum of CT noise, as the spectrum was known to vary across bowtie filter, tube voltage, reconstruction algorithm, and convolution filter. To what extent the noise spectrum impacts the accuracy and precision of NPS measurements remains the effort of our future studies. Second, our phantom-based method evaluates the possible non-linearity of noise and resolution properties in two aspects: the noise magnitude and the object contrast. However, the non-linearity needs to be evaluated for variable anatomical structures also. For example, extremely textured backgrounds (such as lung vessels) and irregular objects (such as speculated lesions) might render a noise and resolution condition not reflected by those measured in uniform backgrounds[23, 58].

Therefore, phantoms with textured background and irregular shaped inserts are warranted in future studies to fully investigate this possible limitation.

In conclusion, this study introduced an experimental method for measuring CT noise and resolution properties relevant to iterative reconstruction and 3D volumetric imaging. Results show that our method is efficient and meaningful in characterizing the system performance accounting for the non-linearity of iterative reconstruction algorithms at multiple contrast and noise levels; accurate and precise to allow challenging measuring conditions with low object contrast and high image noise; and three dimensional to accommodate volumetric operations.

3. Mathematical Prediction of Detection Performance with IR*

3.1 Introduction

The growing number of computed tomography (CT) procedures[59] has raised public concerns regarding the risk associated with ionizing radiation[13]. To reduce CT radiation dose, all major CT vendors have recently introduced commercial iterative reconstruction (IR) algorithms[14-18], which reduce image noise as compared to filtered back-projection (FBP), and therefore potentially provide performance equal to that of FBP at lower acquisition dose[10-12].

To optimize IR for clinical implementation and to compare it with FBP, it is important to evaluate its performance at various imaging conditions with respect to multiple indications. The evaluation, however, is challenged by the fact that IR is likely to be nonlinear, and may have variable noise and resolution depending on the properties of the imaging task[24, 60] (size, contrast, and noise level). As a result, task-independent metrics such as voxel noise or contrast-to-noise ratio are possibly no longer adequate for IR evaluations[10].

Recently, a task-specific metric based on the detectability index (d') has been employed to evaluate IR[25, 44, 46]. The d' mathematically predicts human observer performance according to image and task properties for liner imaging systems[28], and has

* This chapter is based on a manuscript submitted to *Medical Physics*.

been extended to accommodate the nonlinearity of IR images[25, 38, 44, 46]. Previous studies employing *d'* have shown IR's advantage over FBP in terms of dose reduction potentials. However, the studies have been mostly based on a single task[25, 44, 46], while the dose reduction potential of IR might depend on task properties. In addition, to yield a general comparison between IR and FBP that is independent of task, it is necessary to summarize the performance of IR over a wide range of tasks.

The aim of this study was to evaluate IR with both task-specific and task-generic approaches: Task-specific evaluation was based on the performance of IR with respect to specific individual tasks, while the task-generic evaluation was based on the performance of IR summarized across tasks with a wide range of contrast, size, and edge profile to yield a generic assessment. The task-specific performance was evaluated in terms of the task-dependent dose reduction potential of IR, and the task-generic performance was evaluated in terms of the percent of tasks yielding a threshold performance level.

3.2 Methods and Materials

3.2.1 Observer model and detectability index (d')

3.2.1.1 Theory

CT systems were evaluated using an observer model that mathematically predicts human detection performance in terms of a detectability index (*d'*)[27, 28]. *d'* was originally developed for linear systems, but has been extended to accommodate IR

images[25, 38, 46, 61]. There are a host of observer models that could be potentially used for this evaluation[27, 28]. In this study, we deployed non-prewhitening matched filter with eye filter and internal noise (NPWEi) as it has been shown a better match to human observer performance for tasks evaluated in this study[31, 44]. The NPWEi model assumes a signal known exactly scheme with no noise whitening, and further incorporates the sensitivity of human visual system and the inconsistency in human decision-making [27, 28]. It can be calculated as

$$d'^2 = \frac{\left(\iint |W_{task}|^2 \, TTF^2 E^2 dudv\right)^2}{\iint (NPS + N_i)|W_{task}|^2 \, TTF^2 E^4 dudv} \tag{3.1}$$

where u and v are orthogonal spatial frequencies, W_{task} is the task function, TTF is the task transfer function, NPS is the noise power spectrum, E is the eye filter, and N_i is the internal noise. d' can be further used to determine the area under receiver operating characteristic curve (AUC) by assuming Gaussian distributed decision variable[27], as

$$AUC_{d'} = \frac{1}{2}[1 + erf(\frac{d'_{adj}}{2})]. \tag{3.2}$$

Task function characterizes the profile, size, and contrast of the task. It is mathematically modeled to be the Fourier-difference between the null hypothesis, "absent", and the alternative hypothesis, "present", as

$$W_{task}(u,v) = \left|\mathcal{F}_{2D}\{H_1(x,y) - H_0(x,y)\}\right| = \left|\mathcal{F}_{2D}\{T(x,y)\}\right|, \tag{3.3}$$

where H_0 is the null hypothesis, H_1 is the alternative hypothesis, and T is the detection task (the lesion) mathematically defined in spatial domain.

TTF and NPS characterize the image quality. TTF is a task specific extension of the concept modulation transfer function (MTF) to accommodate the potential nonlinearity of iterative reconstruction algorithms[24, 38, 46, 61]. TTF characterizes the resolution of the image with respect to a given "task", i.e., an object to be detected with specific contrast and background noise level, such that even if the system behaves nonlinearly, the TTF will still characterize the transformation of the task to the image; it is assumed that the system's resolution varies sufficiently constant in the contrast range of the task that can be considered constant for the given task. The TTF is calculated as the Fourier transform of the line spread function (LSF) with respect to a specific object contrast and an image noise magnitude.

NPS reflects the magnitude and texture of the quantum noise. It is calculated as the square of the noise (variance) in the Fourier domain as a function of spatial frequency. Traditionally, NPS is defined independent of the noise magnitude. In this study, to capture the possible change in the shape of the NPS with nonlinear IR algorithms, the NPS was assumed changeable as a function of the magnitude of the noise. As such, both the TTF and the NPS characterized the properties of the system specific to the targeted task and background.

Both eye filter and internal noise characterize the properties of the observer. Eye filter characterizes the sensitivity of human visual system to various spatial frequencies[62, 63], which is mathematically modeled as

$$E(u,v) = \left(u^2 + v^2\right)^{0.65} \exp\left(-c\left(u^2 + v^2\right)\right), \tag{3.4}$$

where u and v are orthogonal spatial frequencies and c is a scale factor depending on the viewing distance and display size, selected to yield peak sensitivity at 4 cycles/degree. Internal noise reflects the cognitive decision inconsistency of the observer. The idea is that even with identical detection task, inconsistent decisions may be reached in repeated trials. Previous studies have shown that the internal noise impacts the detection performance in a way similar to a 60% increase in the quantum noise[28, 32].

3.2.1.2 Experimental setup

The TTF and NPS were experimentally measured from a previously developed image quality phantom[38] (Figure 3.1). The phantom had four sections of increasing diameters to represent patients of increasing sizes. In this study, both NPS and TTF were measured from the 30-cm section (29-cm water equivalent diameter) to represent the image quality when imaging a medium sized patient. The TTF was measured from the rod inserts of various attenuations using an edge technique, and the NPS was measured from the uniform region of the phantom. The measurement methods for the TTF and NPS have been validated to ensure <10% errors and described in detail in Chapter 2 and in previous publications[38, 61, 64].

The phantom was scanned on a CT scanner (SOMATOM Definition Flash, Siemens Healthcare, Forchheim, Germany) at 120 kVp with five radiation exposure levels (volume CT dose index, $CTDI_{vol}$ = 64.8, 27.2, 13.6, 6.9, and 3.4 mGy). The data was reconstructed with FBP and IR (Iterative Reconstruction in Imaging Space, IRIS; Sinogram Affirmed Iterative Reconstruction with strengths of 3 and 5, SAFIRE3 and SAFIRE5). The TTF was calculated with respect to the noise level, the contrast level, and the reconstruction algorithm. The NPS was calculated with respect to the noise level and the reconstruction algorithm.

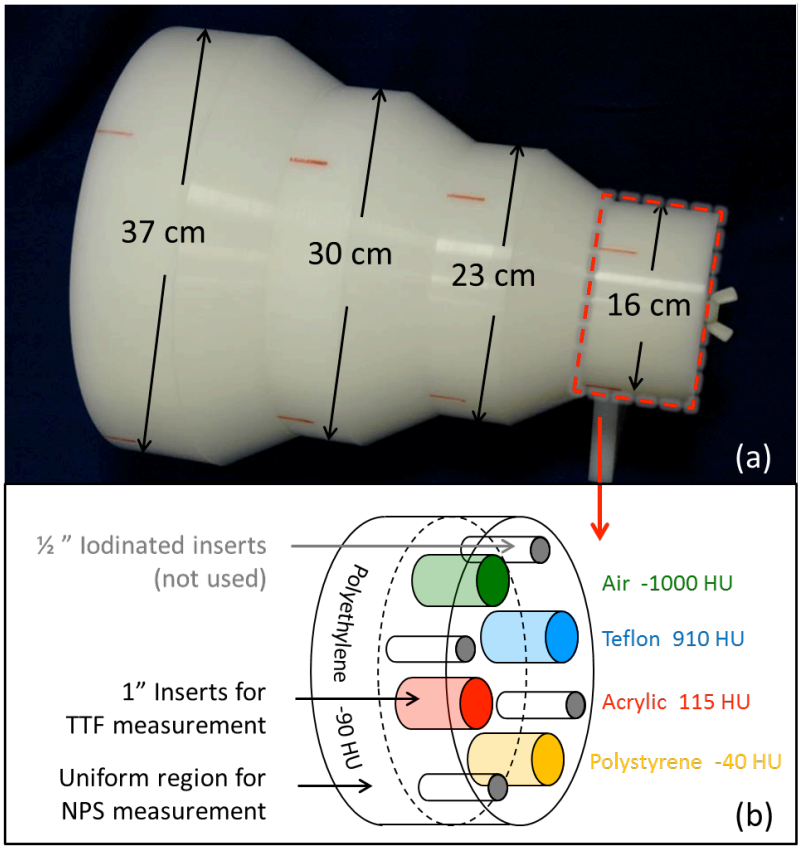

Figure 3.1: (a) The phantom for TTF and NPS measurements is composed of four cylindrical sections of increasing diameters, with tapered sections in between. (b) Each cylindrical section is divided into two subsections, one with uniform regions for the NPS measurement and the other with rod inserts of varying contrasts for the TTF measurement[38].

In this study, the task was mathematically defined in spatial domain using a previously developed contrast profile[65-67], as

$$C(r) = C_{peak}\left(1-\left(\frac{r}{R}\right)^2\right)^n, \tag{3.5}$$

where C_{peak} is the peak contrast, r is the radial distance from the center, R is the radius, and n is a positive number inversely related to the steepness of the contrast profile. In addition, a second radius R_{20} was defined at 20% of the peak contrast to approximate the radius visible to human observers[65], calculated as

$$R_{20} = R\sqrt{1-\sqrt[n]{0.2}}\,. \tag{3.6}$$

The contrast profile was then employed to simulate a library of tasks, including tasks of multiple sizes (R_{20} from 1 to 4 mm, 50 intervals), contrast levels (C_{peak} from 10 to 100 HU, 50 intervals), and edge profiles (sharp edge, n = 1; and soft edge, n = 3). Figure 3.2(a) shows selected tasks of 2 contrast levels, 3 sizes, and 2 edge profiles. Note that the tasks of different edge profiles share the same R_{20}. The library of tasks was then Fourier transformed into a library of task functions, with examples shown in Figure 3.2(b). The intensity (power) of the task functions depends on the task contrast and size, and the

frequency component (shape) of the task function depends on the task size and edge profile.

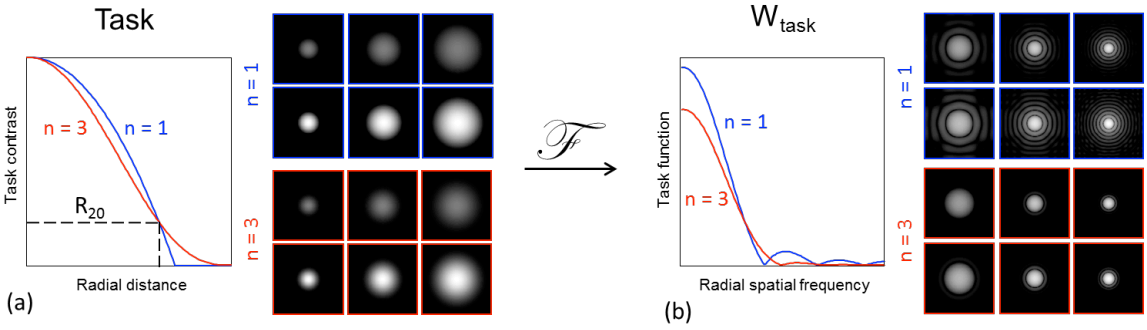

Figure 3.2: (a) Examples of tasks defined at 3 sizes, 2 contrast levels, and 2 edge profiles (n=1 and n=3). (b) The corresponding task functions.

For the calculation of the eye filter, the viewing distance was assumed to be 100 cm, and the display size was assumed to be 40 cm. The corresponding c was calculated to be 12.4 mm2 by differentiating Eq. 3.4. The internal noise was modeled to be 60% of the quantum noise to account for the cognitive inconsistency. In addition, because in clinical practices the unknown task location caused additional deterioration to observer performance[32], the *d'* values calculated with Eq. 3.1 were adjusted with an offset factor of 1.5 according to a previous observer study[46].

3.2.2 Task-specific and task-generic evaluations

As the observer performance ($AUC_{d'}$) improves monotonically as dose increases, a minimum dose requirement (the threshold dose level) can be defined to achieve a threshold AUC level. A threshold AUC of 0.9 was chosen in this study to provide a good

tradeoff between diagnostic accuracy and patient dose. The dose reduction potential of IR was then calculated as the difference between IR and FBP's threshold dose levels for specific tasks. A relative dose reduction potential was also calculated as the ratio of the dose reduction potential of IR and the threshold dose level of FBP.

In addition, the task-generic performance of IR was evaluated in terms of the percentage of tasks (out of the entire library developed in 3.2.1.2) yielding a performance higher than that defined by the threshold AUC level. This metric did not focus on the performance with a specific task, but rather the general performance of the system; the larger the percentage, the better the system performance in general.

3.3 Results

3.3.1 TTF and NPS

Figure 3.3 shows examples of phantom-measured TTF as a function of reconstruction algorithm and object contrast. Compared to FBP, IRIS had similar TTF at all contrast levels, while SAFIRE3 and SAFIRE5 had enhanced TTF at low contrast levels (50% frequency of polystyrene-based TTF: FBP, 0.41 mm^{-1}; IRIS, 0.42 mm^{-1}; SAFIRE3, 0.44 mm^{-1}; and SAFIRE5, 0.46 mm^{-1}). All TTF had peak values slightly off zero frequency, as a result of the edge enhancement employed in the reconstruction algorithm.

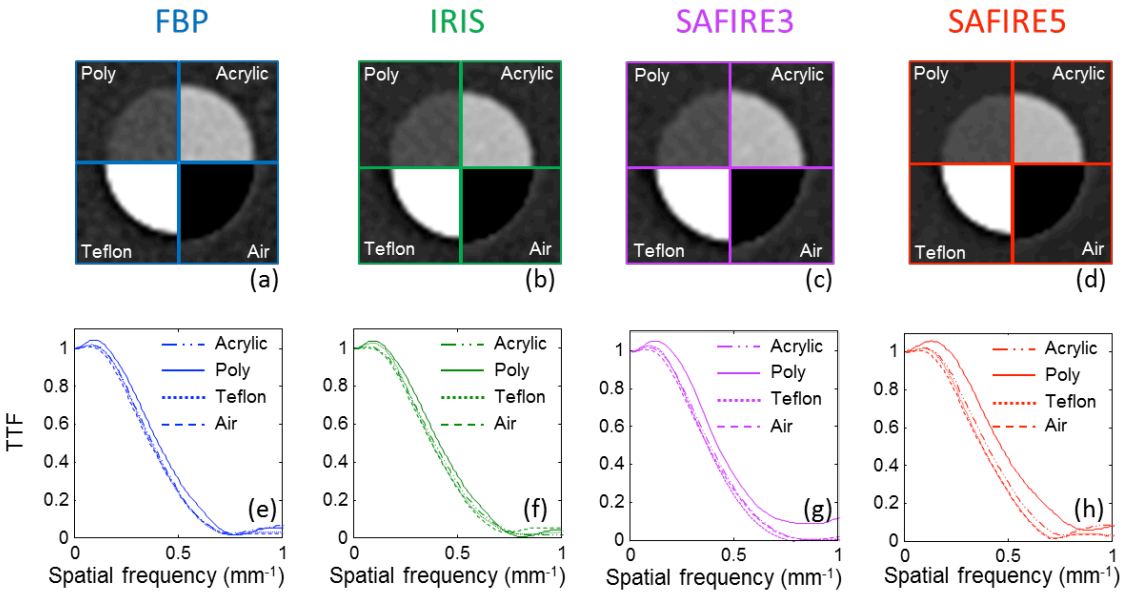

Figure 3.3: (a,b,c,d) The regions of interest showing the rod inserts of varying contrasts and (e,f,g,h) the corresponding TTF curves.

Figure 3.4 shows the NPS results as a function of reconstruction algorithm and imaging dose level. Compared to FBP at the same dose level, IR showed notable reductions in the NPS magnitude: 50%, 50%, and 80% for IRIS, SAFIRE3, and SAFIRE5, respectively. Because the NPS magnitude is proportional to the square of the noise magnitude, the corresponding reductions in noise magnitude were: 30%, 30%, and 55% for IRIS, SAFIRE3, and SAFIRE5, respectively. Another difference with IR was the reduced NPS peak frequency: IR, especially SAFIRE5, had their peaks shifted to lower frequencies, reflecting IR's "waxier" texture (peak locations: FBP, 0.25 mm^{-1}; IRIS, 0.23 mm^{-1}; SAFIRE3, 0.23 mm^{-1}; and SAFIRE5, 0.16 mm^{-1}).

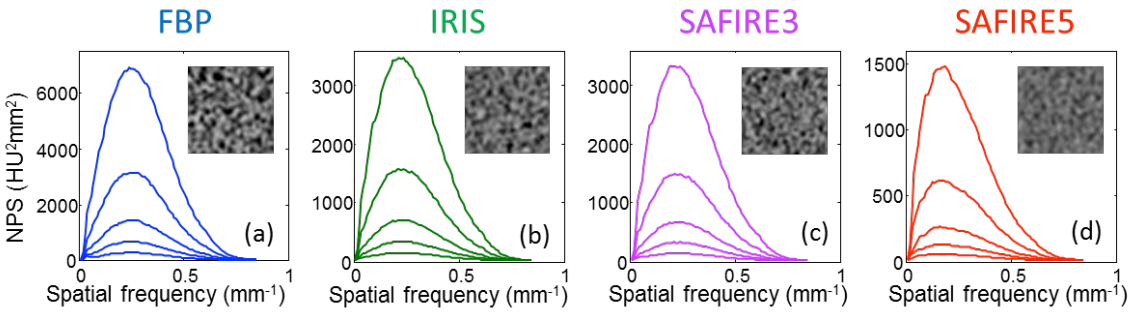

Figure 3.4: The NPS results measured at five dose levels for (a) FBP, (b) IRIS, (c) SAFIRE3, and (d) SAFIRE5. Examples of regions of interest for NPS measurements are also shown.

3.3.2 Detectability index

Examples of d' results for FBP and IR are plotted in Figure 3.5 as a function of task size, contrast, and edge profile, with the dose level fixed at 13.6 mGy. For the same task, SAFIRE5 yielded the highest d' (best performance), followed by SAFIRE3 and IRIS, and FBP had the lowest d'. For a fixed reconstruction algorithm, larger task size, higher task contrast, and sharper task edge profile led to higher d'.

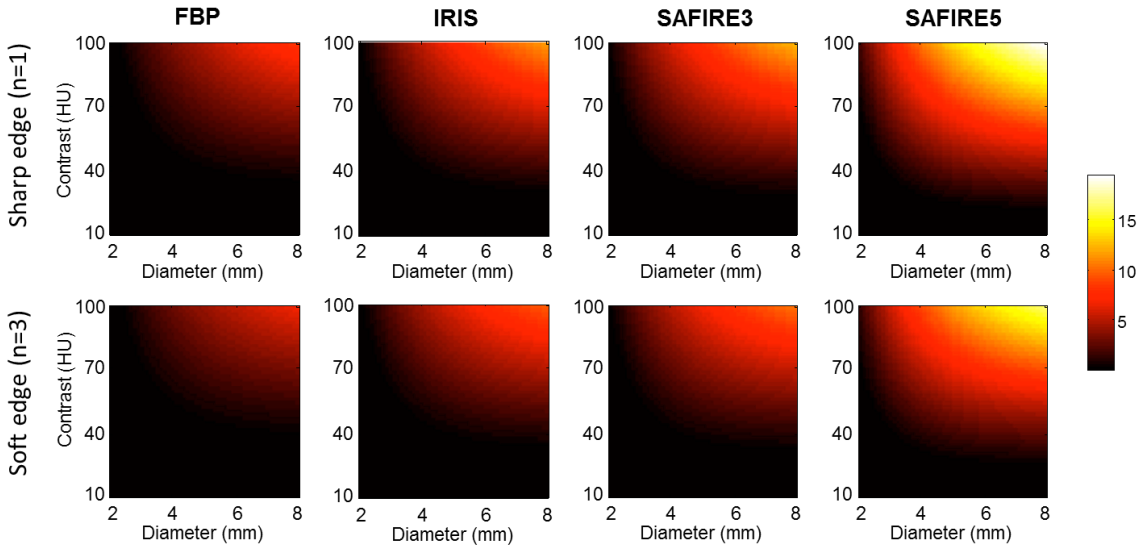

Figure 3.5: The detectability index (d') for four reconstruction algorithms, plotted as a function of task size, task contrast, and task edge profile.

3.3.3 Task-specific performance: dose reduction potential

The dose reduction potentials are summarized in **Table 3.1**. In general, SAFIRE5 showed the most significant dose reduction potentials (11-54 mGy), followed by SAFIRE3 (7-36 mGy) and IRIS (6-26 mGy). The dose reduction potentials highly depended on task size and task contrast, with tasks of lower contrasts and smaller sizes, i.e., more challenging tasks, showed higher dose reduction potentials. Softer edge profile showed higher dose reduction potentials with SAFIRE3 and SAFIRE5, but not with IRIS. The relative dose reduction potentials generally followed the same trend, but were less dependent on task properties (IRIS, 37-50%; SAFIRE3, 50-61%; and SAFIRE5, 77-84%).

Table 3.1: The dose reduction potentials of IR in the unit of mGy. The numbers in parentheses are relative dose reduction potentials.

R_{20} (mm)	C_{peak} (HU)	Sharp edge profile (n = 1)			Soft edge profile (n = 3)		
		IRIS	SAFIRE3	SAFIRE5	IRIS	SAFIRE3	SAFIRE5
1	100	25 (50%)	30 (61%)	42 (84%)	26 (44%)	35 (59%)	50 (84%)
1.5	100	7 (41%)	8 (52%)	13 (79%)	8 (41%)	11 (53%)	16 (80%)
1.5	65	25 (47%)	32 (60%)	45 (84%)	24 (37%)	36 (56%)	54 (83%)
2.2	65	6 (40%)	7 (50%)	11 (77%)	8 (40%)	10 (52%)	15 (79%)
2.2	45	24 (48%)	30 (60%)	41 (83%)	24 (37%)	35 (55%)	53 (82%)
3	45	10 (42%)	13 (52%)	19 (78%)	13 (50%)	24 (59%)	33 (81%)
4	45	8 (41%)	9 (50%)	15 (77%)	15 (47%)	17 (54%)	24 (78%)

An example of the dose reduction calculation is shown in **Figure 3.6** for a task at 2.5-mm in radius (R_{20}), 54-HU peak contrast, and soft edge profile (n=3). The $AUC_{d'}$ values for FBP and SAFIRE3 are plotted as a function of dose level. Simulated images showing the task in noisy backgrounds are displayed next to each $AUC_{d'}$ data point for demonstration. The threshold doses required to achieve the threshold AUC were 13 and 28 mGy for SAFIRE3 and FBP, respectively. The dose reduction potential of SAFIRE3 for this task was therefore calculated to be 28-13 = 15 mGy. The relative dose reduction potential of SAFIRE3 was calculated as 15/28 = 54%. Note that for more conspicuous tasks (high contrast, large diameter, or sharp edge) or more obscure tasks (low contrast, small diameter, or soft edge), the $AUC_{d'}$ curves might not cross the horizontal line of the threshold AUC, not yielding a meaningful interpolation of the dose reduction potential.

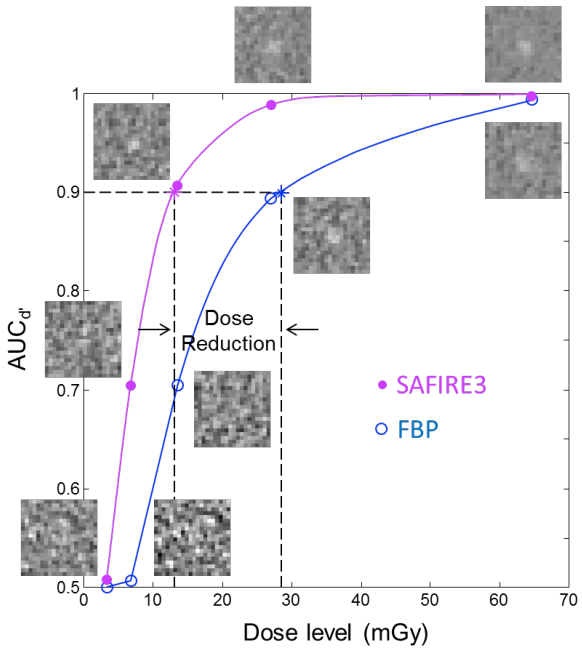

Figure 3.6: The calculation of the dose reduction potential of SAFIRE3 for a task with 2.5-mm R_{20}, 54-HU peak contrast, and n=3 (soft) edge profile.

3.3.4 Task-generic performance: percent of tasks exceeding threshold AUC

Figure 3.7 plots the tasks yielding the threshold AUC (0.9) as lines across the task space. Tasks above each line had AUC exceeding the threshold. The percent of tasks exceeding the threshold AUC, visually interpreted as the ratio between the area above each line and the area of the entire task space, is numerically summarized in Table 3.2. In general, IR allowed more tasks to exceed the threshold AUC than FBP, with an increase of 8-12% with IRIS, 10-16% with SAFIRE3, and 20-33% with SAFIRE5 (note: although percentage signs are used here, these are absolute differences). Higher dose level and sharper edge profile also allowed more tasks to achieve the threshold performance, as

expected. However, the advantage of IR was generally less apparent at higher dose levels.

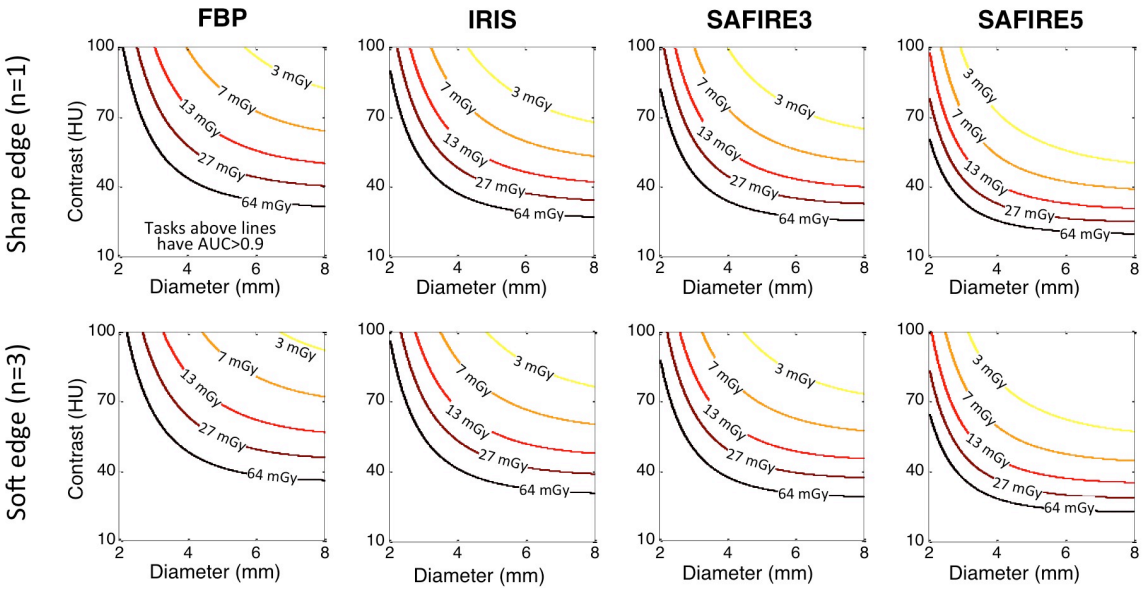

Figure 3.7: Tasks yielding the threshold AUC of 0.9 at different dose levels, specified as lines of different colors.

Table 3.2: The percent of tasks exceeding the threshold AUC, calculated for 2 edge profiles, 5 dose levels, and 4 reconstruction algorithms.

Dose level	Sharp edge profile (n = 1)				Soft edge profile (n = 3)			
	FBP	IRIS	SAFIRE3	SAFIRE5	FBP	IRIS	SAFIRE3	SAFIRE5
3 mGy	5	14	17	35	1	9	11	28
7 mGy	18	30	34	51	12	24	28	45
13 mGy	34	45	49	64	28	39	43	59
27 mGy	47	57	60	73	42	51	55	69
64 mGy	61	69	72	81	56	64	68	78

3.4 Discussion

The evaluation of IR needs to be task-specific to capture the potential dependency of IR performance on the task, but also task-generic to enable a general comparison of IR and FBP. In this study, we accommodated both needs by evaluating IR with tasks across a range of sizes, contrasts, and edge profiles. Results showed that IR's dose reduction potential depended on the size, contrast, and edge-profile of the task. Results also showed that across a wide range of tasks, IR was superior to FBP in terms of the percent of tasks exceeding a threshold performance level.

In previous studies, IR was mostly assessed with task-independent metrics, such as voxel noise and contrast-to-noise ratio[10, 68, 69]. As a result, a single dose reduction potential independent of tasks was usually prescribed to IR. Our task-specific analysis, however, demonstrated that tasks with lower contrast and smaller size had higher absolute dose reduction potential (differences up to 20, 29 and 43 mGy for IRIS, SAFIRE3, and SAFIRE5, respectively). This highlights the importance of evaluating IR with respect to the task, and could be employed to optimize the operating dose of IR according to the size and contrast of the task of interest.

The task-generic analysis evaluated overall performance and enabled general comparisons between FBP and IR. The framework can be further used to assess the performance of IR for a specific medical indication. For example, by summarizing over

tasks with kidney stone's typical contrast and size levels, the general detectability of kidney stones with IR can be assessed.

In this study, in order to calculate Fourier-based d' for IR, the Fourier concepts of MTF and NPS were extended to contrast-dependent TTF and magnitude dependent-NPS to accommodate IR's potential nonlinearity. These extensions are not necessary if d' is calculated in the spatial domain[57]. However, the spatial domain approach is practically limited in that it can be considerably time-consuming: Hundreds of repeated scans are required to characterize the covariance matrices for a single task, let alone that a range of tasks need to be evaluated to fully examine IR. While approximating, the Fourier approach extended to accommodate certain nonlinearity, and agrees well with human observer performance across three IR algorithms from three CT vendors[44, 46].

The threshold AUC of 0.9 was employed for the evaluation of IR, because it provided a good tradeoff between diagnostic accuracy and patient dose. However, one question is that how will the choice of the threshold AUC affect the assessment. To answer this question, the percent of tasks exceeding the threshold AUC were recalculated with a lower threshold AUC of 0.7, as shown in **Table 3.3**. Compared to Table 3.2, the percentage of tasks yielding the lower threshold AUC is higher, especially at lower dose levels. However, the advantage of IR in terms of the difference between FBP and IR remained largely unchanged (6-11% for IRIS, 9-15% for SAFIRE3, and 9-31% for SAFIRE5).

Table 3.3: The percent of tasks yielding an AUC higher than 0.7.

Dose level	Sharp edge profile (n = 1)				Soft edge profile (n = 3)			
	FBP	IRIS	SAFIRE3	SAFIRE5	FBP	IRIS	SAFIRE3	SAFIRE5
3 mGy	19	30	34	50	13	24	27	44
7 mGy	34	44	48	63	28	39	43	58
13 mGy	48	57	61	74	42	52	56	70
27 mGy	60	67	71	81	55	63	66	77
64 mGy	71	77	80	87	67	73	76	85

In this study, both dose reduction potential and detectable range were calculated for detection tasks only. Therefore, the advantage of IR derived in this study may not be translated to other types of tasks, such as discrimination tasks that distinguish two features, and quantification tasks that quantify the physical properties of certain features. For example, CT data has been employed to quantify the development of lung nodule volume over time. For such task, two commercial IR algorithms showed little advantage in terms of quantification precision[23]. This once again highlights the importance of evaluating IR with respect to the imaging task at hand.

The NPS of IR was measured from uniform background, which, due to IR's nonlinearity, is not representative of the noise property in heavily textured anatomical background, such as lung[23, 58]. Furthermore, the d' model employed in this study only accounted for the impact of quantum noise, not anatomical noise. Therefore, the evaluation of IR in this study is more applicable to tasks in relatively uniform

background, such as hepatic lesions and kidney stones. Our ongoing efforts are focusing on measuring the quantum and anatomical NPS from textured backgrounds, such that IR can be evaluated for tasks in textured backgrounds.

To conclude, this study went beyond traditional contrast and noise based assessments of IR, and developed a framework for IR evaluations with respect to the detection task. The task-specific evaluation demonstrated the strong dependency of IR's dose reduction potential on task size and contrast, which can be employed to optimize IR for clinical indications with specific ranges of size and contrast. The task-generic evaluation demonstrated IR's overall superiority over FBP in terms of the range of the tasks detectable, which can be employed for general comparisons between algorithms.

4. Experimental Assessment of Quantification Performance with IR*

4.1 Introduction

Size estimations of pulmonary nodules provide useful quantitative information for diagnosis and staging of lung cancer. Currently, these estimations are standardized by uni-dimensional (Response Evaluation Criteria in Solid Tumors, RECIST) and bi-dimensional (World Health Organization, WHO) methods[70, 71], which merely measure the diameters of the nodule by assuming symmetrical nodule shrinkage/growth. This assumption of symmetrical development, however, is not always valid. For a more comprehensive estimation, new three-dimensional (3D) methods have been proposed to estimate the size in terms of nodule volume. The accuracy (bias) and precision (variance) of this 3D method, however, are subject to many factors such as scanning and reconstruction parameters, segmentation software, and nodules characteristics. A detailed knowledge of the accuracy and precision is required to confidently estimate volume changes[33].

In addition, iterative reconstruction algorithms (IR) recently introduced in the clinics have greatly improved image contrast-to-noise ratio and demonstrated significant potential advantage in interpretive imaging as compared to conventional filtered back-projection (FBP) images[19, 69, 72]. The effect of IR algorithms on quantitative imaging,

* This chapter is based on a manuscript published on *Medical Physics,* titled "Volumetric quantification of lung nodules in CT with iterative reconstruction (ASiR and MBIR)."

however, has not received much attention. Specifically, two IR algorithms of Adaptive Statistical Iterative Reconstruction (ASiR) and Model Based Iterative Reconstruction (MBIR, commercially marketed as Veo) notably reduce image noise by selectively identifying and subtracting noise from images[12, 20, 69], potentially enabling more accurate and precise volume quantifications as compared to FBP. MBIR further improves image resolution and suppresses certain artifacts by modeling system-optics, noise, and non-idealities of the data[12, 26]. One may assume such attributes to help improve the estimation of nodule volumes, an assumption that needs verification.

The purpose of this study was to investigate the impact of iterative reconstruction on the accuracy and precision of lung nodule volume quantifications. In addition, this impact was evaluated as a function of other key influencing parameters, including dose and reconstruction slice thickness.

4.2 Method and Materials

4.2.1 Image acquisition and reconstruction

An anthropomorphic chest phantom consisting of realistic pulmonary vessels (LUNGMAN, KYOTO KAGAKU, Kyoto, Japan) was used to simulate chest CT imaging (Figure 4.1). The phantom was supplemented with pulmonary nodules simulated by acrylic spheres (80 HU at 120 kVp) of diameters 4.76 and 9.53 mm, attached to its vessels and pleura. A total of 14 small and 21 large nodules were placed in various locations throughout the lung structure. Compared to uniform chest phantoms employed in previous study[73-75], the vessel structures in our phantom were expected to provide more realistic anatomy and geometry.

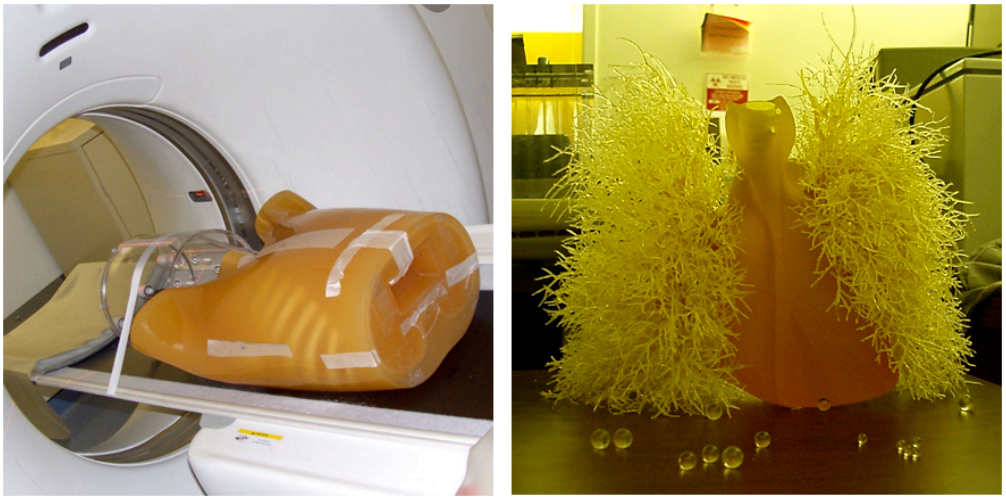

Figure 4.1: (a) An anthropomorphic chest phantom was used to simulate chest CT images, which contained (b) a lung insert with realistic pulmonary vessels. Acrylic spheres of diameters 4.76 and 9.53 mm were attached to the vessels and pleura to simulate pulmonary nodules.

The phantom was scanned on a Discovery CT750 HD CT scanner (GE Healthcare, Waukesha, WI) with a standard chest protocol at Duke University Medical Center: 40 mm collimation, 120 kVp, 1.375 pitch, 0.5 s rotation time, large bow tie filter, 500-mm scan field of view, 360-mm reconstruction field of view, a clinical dose level corresponding to a noise index of 28.0 ($CTDI_{vol} \approx 7.5$ mGy) with automatic exposure control (AEC) turned off, filtered back-projection (FBP) reconstruction with "detail" kernel, and 1.25 mm slice thickness with equal slice spacing. In addition to the routine protocol, two additional iterative reconstructions, ASiR (blended with 50% of FBP to moderate the "patchy" texture of noise in IR images) and MBIR, were used. Two other important parameters, dose and slice thickness, were also expanded to a broader parameter space to examine their relative impact. Added data included five additional dose levels of 75%, 50%, 25%, 10%, and 3% (100% corresponding to the standard dose level), and two additional slice thicknesses of 0.625 and 2.5 mm. AEC was turned off at all dose levels. The slice spacing was equal to the slice thickness for all slice thicknesses. As illustrated in Figure 4.2, in general, this study investigated three reconstruction algorithms, six dose levels, and three slice thicknesses, yielding fifty-four distinct acquisition and reconstruction protocols. Furthermore, each protocol was repeated five times to enable the estimation of the precision of quantifications.

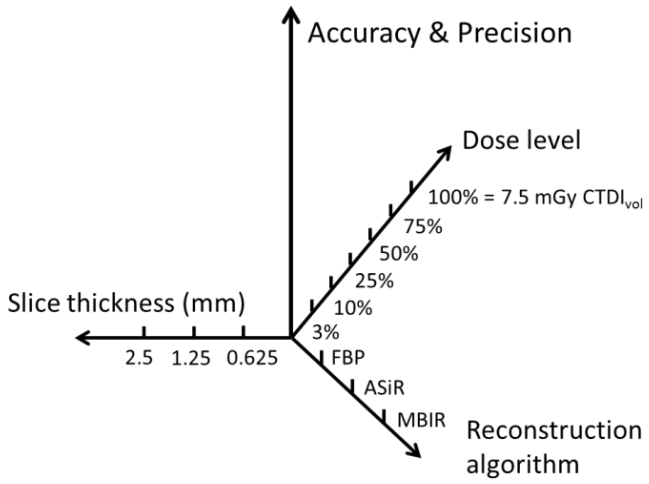

Figure 4.2: One routine chest protocol in Duke University Health Center was expanded into fifty-four protocol to investigate the impact of protocol parameters on the accuracy and precision of volume quantifications under various conditions, including three reconstruction algorithms (FBP, ASiR, and MBIR), six dose levels (3%, 10%, 25%, 50%, 75%, and 100%, with 100% ≈ 7.5 mGy CTDIvol), and three slice thicknesses (0.625, 1.25, and 2.5 mm).

4.2.2 Data analysis

As a related factor to volume quantifications, the image noise magnitude was quantified from the uniform regions in the mediastinum of the anthropomorphic phantom with a 50 x 50 ROI (35 mm x 35 mm). The image appearance in terms of resolution and noise texture was further visually compared between FBP, ASiR, and MBIR images reconstructed from the same raw data.

Nodule volumes were quantified with clinical software Lung VCAR (GE Healthcare, Waukesha, WI; hereinafter referred to as Software A) in a semi-automatic fashion[34]. After loading the CT dataset, the software first suggest all potential nodules

with its built-in computer-aided detection feature, as highlighted in Figure 4.3 (a,b). The user then can click on the center of a nodule (whether suggested by the software or not) to initiate the segmentation. The software automatically choose a built-in segmentation algorithm that best matches the nodule's shape, density and surrounding structure, and reports the segmented volume in mm3 [Figure 4.3 (c,d)]. All 9450 nodule image dataset (54 protocols x 5 repetitions x 35 nodules) were segmented using this procedure.

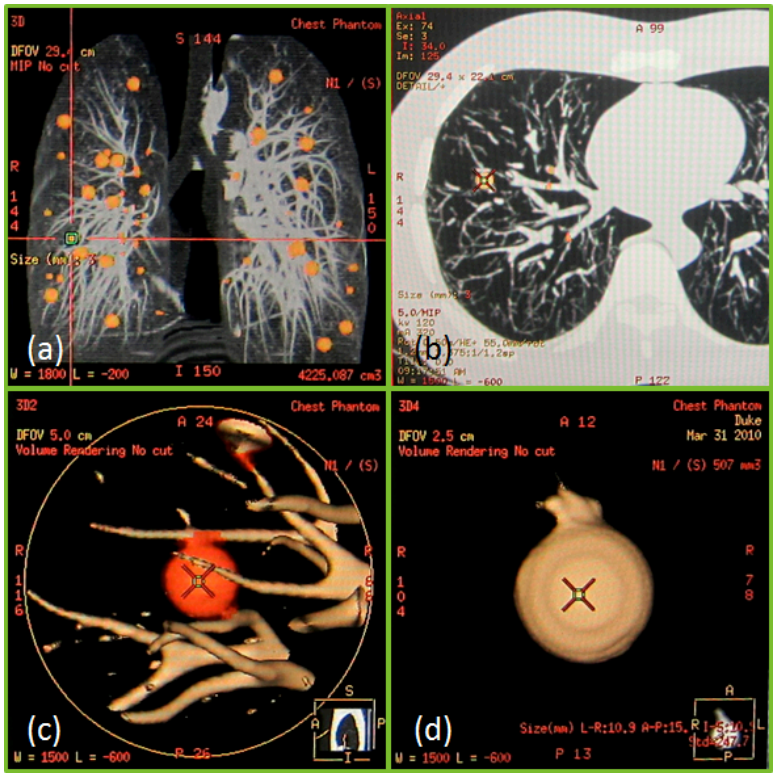

Figure 4.3: Nodule volumes were semi-automatically quantified from the images using clinical software Lung VCAR. (a) Coronal view, nodules suggested for quantifications are highlighted in circles, and the nodule selected for quantification is highlighted at the crossing. (b) Axial view, the nodule selected for quantification is highlighted at the crossing (c) 3D rendering of the selected nodule with surrounding structures (d) 3D rendering of the segmented nodule with volume reported in mm^3.

Since the segmentation software is a large factor of whether IR improves or degrades volume quantification, a second clinical software, iNtuition (TeraRecon, Foster City, CA; hereinafter referred to as Software B), was employed for volume quantifications. Similar to Software A, Software B is semi-automatic in the sense that it relies on the users to input the location of the nodule. The segmentation process of Software B, however, is more generic than software A, as it is developed for processing images from multiple vendors. The segmentation process of software B further allows the user to choose the type of segmentation (single click, region growing, or manual ROI) and to manually modify the boundary of the segmented nodule. Most segmentation procedures were performed with "single click" with no manual modification to the nodule boundary. Only two 4.76 mm nodules were segmented with "region growing", as they were attached to the pleura and could not be properly segmented with "single click". All 9450 nodule image dataset (54 protocols x 5 repetitions x 35 nodules) that segmented with software A were also segmented with software B.

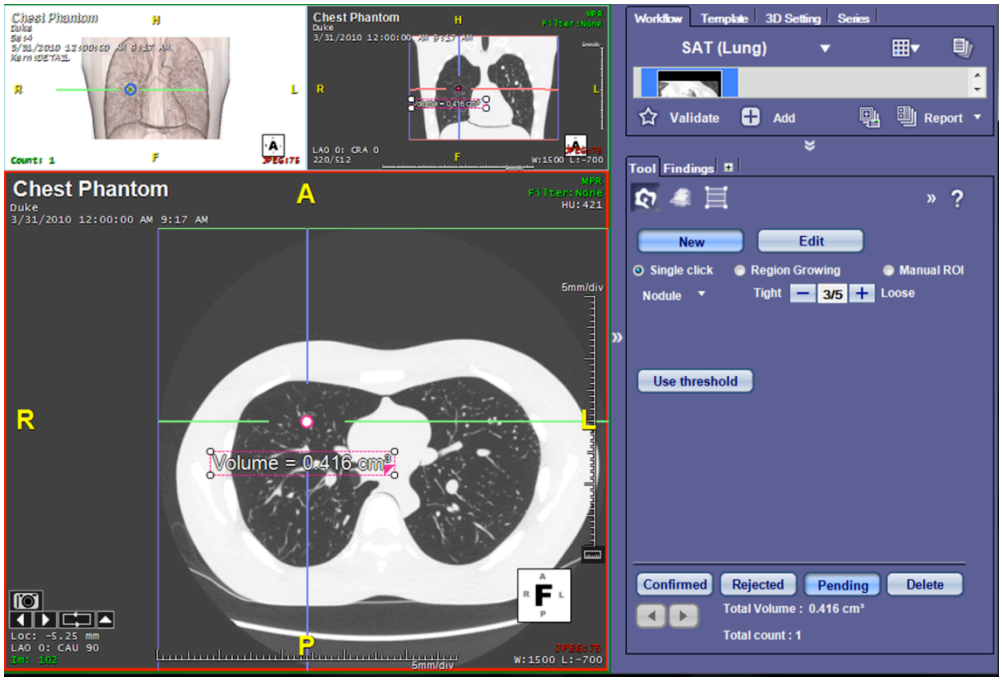

Figure 4.4: A second clinical software, iNtuition, was employed for nodule segmentations. The user inputs the location of the nodule and the type of segmentation. The software reports the volume in cm^3.

Following the volume quantification, each protocol's accuracy and precision were calculated in terms of bias and repeatability coefficient, respectively[76]. Bias was the difference between the segmented nodule volume and the true volume. A bias with a smaller absolute value shows a measurement closer to the truth, thus a better accuracy. For jth protocol, which included n repetitions and K nodules, bias was calculated as

$$B_j = \frac{1}{nK}\sum_{i=1}^{n}\sum_{k=1}^{K}(V_{ijk} - V_t), \tag{4.1}$$

where Vt and Vijk represent the true and measured volume, respectively. The true volume was calculated based on the diameter of the spherical nodule, and was

physically validated using a liquid displacement method. The percent bias (PB) was also calculated, as

$$PB_j = \frac{B_j}{V_t} \times 100. \quad (4.2)$$

The repeatability coefficient (RC) was defined as the expected absolute difference between any two repeated quantifications of the same object, for 95% of cases. Therefore, a smaller RC reflects a more precise measurement. For jth protocol, RC was calculated as

$$RC_j = 1.96\sqrt{2\sigma_{Wj}^2} = 2.77\sigma_{Wj} \approx 2.77\hat{\sigma}_{Wj}, \quad (4.3)$$

where

$$\hat{\sigma}_{Wj} = \sqrt{WMS_j} = \sqrt{\sum_{i=1}^{n} WMS_{ij} / n} = \sqrt{\sum_{i=1}^{n} \sum_{k=1}^{K} \frac{(V_{ijk} - \bar{V}_{ij})^2}{n(K-1)}}. \quad (4.4)$$

WMS_j was the estimate for the within-nodule variance (σ_{Wj}^2), obtained from one-way analysis of variance (ANOVA) with the nodule location as the main factor. RC reflected the absolute fluctuation for each volume measurement. RC was further normalized by the true nodule volume to yield the percent repeatability coefficient (PRC), as

$$PRC_j = \frac{RC_j}{V_t} \times 100. \quad (4.5)$$

Comparisons of accuracy and precision difference between protocols were carried out via a generalized estimating equation (GEE) approach[76, 77]. The GEE analysis accounts for the correlation of multiple measurements from the same nodule. Since

previous studies has demonstrated the impact of dose and slice thickness with FBP reconstruction, the goal of this study was focused more primarily on the comparisons of the two new iterative reconstructions versus FBP, at various dose levels and slice thicknesses. A p-value less than or equal to 0.05 was considered to be statistically significant with no adjustment for multiple comparisons. A p-value greater than 0.05 but less than 0.1 was considered to be marginally significant. All analyses were carried out by a standard statistical package (SAS version 9.2, SAS Institute Inc., Cary, NC).

4.3 Results

All twenty-one 9.53 mm nodules were successfully segmented. Two of the fourteen 4.76 mm nodules failed in the segmentation of Software A with images scanned at 25%-100% dose levels and reconstructed with MBIR. The nodules for which the segmentation failed were highly cluttered with surrounding vascular structures. Only the successfully segmented nodules were analyzed for accuracy and precision.

4.3.1 Image noise and appearance

The noise magnitude measured from the uniform region of the phantom is shown in **Table 4.1**. Higher dose and thicker slice thickness led to less noise. At fixed dose level and slice thickness, ASiR reduced the noise magnitude by ~30% throughout all dose levels as compared to FBP, and MBIR reduced the noise magnitude by ~35% at 100% dose level and ~80% at 3% dose level. The noise magnitude, however, did not fully capture the difference between FBP and IR images. As shown in Figure 4.5, the noise

texture and the resolution were also different between FBP and IR, which are especially pronounced at low dose levels in the soft tissue.

Table 4.1: The noise measured from the uniform region of the phantom image.

Reconstruction algorithm		**FBP**			**ASiR**			**MBIR**		
Slice thickness (mm)		**0.625**	**1.25**	**2.5**	**0.625**	**1.25**	**2.5**	**0.625**	**1.25**	**2.5**
Dose levels	**100%**	13.0	11.1	7.7	9.1	7.8	5.5	8.2	7.4	6.3
	75%	14.9	12.9	8.9	10.6	9.1	6.3	8.7	8.0	6.7
	50%	18.1	15.0	10.4	12.6	10.3	7.2	9.1	8.3	6.8
	25%	25.3	21.5	14.9	17.8	15.0	10.3	9.8	9.2	7.8
	10%	41.7	36.2	25.5	28.8	25.4	18.1	11.7	10.8	9.5
	3%	74.8	63.6	43.8	52.5	45.1	30.4	15.3	14.3	12.6

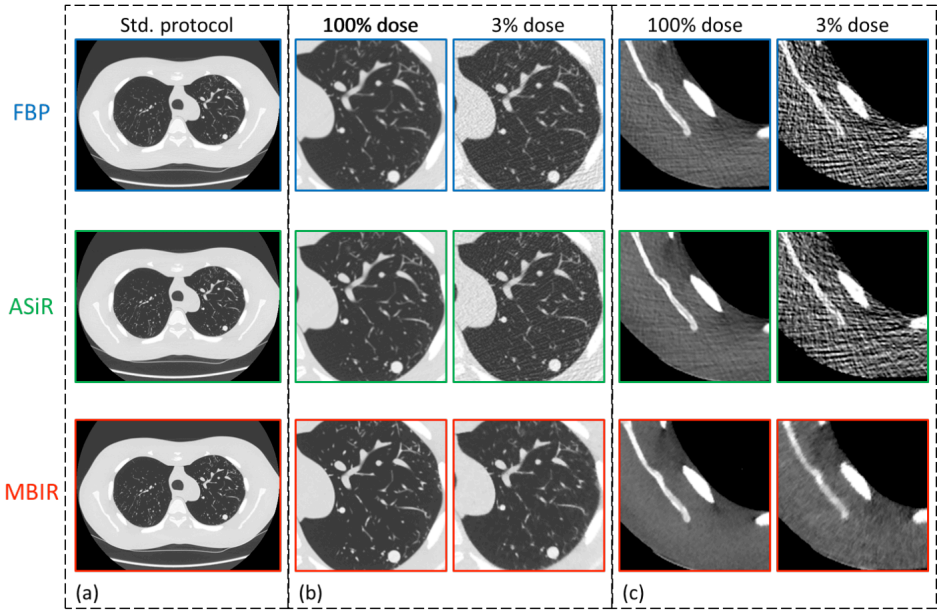

Figure 4.5: Sample images of the phantom reconstructed with FBP, ASiR, and MBIR. (a) The phantom imaged with the standard chest protocol (100% dose level = 7.5 mGy, 120 kVp, 1.25 mm slice thickness). (b) Magnified views of lung region with lung window (WL/WW = -600/1600). (c) Magnified views of soft tissue with soft tissue window (WL/WW = 40/400).

4.3.2 Accuracy and precision of volume quantifications – Software A

In order to efficiently navigate through the large amount of data generated in this study, we first demonstrate the accuracy and precision of volume quantifications with Software A. The impact of IR on the accuracy and precision of 9.53 mm nodules is shown in Figure 4.6. Figure 4.6 (a) shows the accuracy of the volume quantified with 2.5 mm slice thickness. At 100% dose level (7.5 mGy), which corresponds to the clinical dose level used in our institute, FBP and MBIR show comparable accuracy. ASiR, on the other hand, shows a higher PB, i.e., a worse accuracy. This difference between ASiR and the

other two algorithms is consistently observed throughout all other five dose levels (3%-75%, 0.23 – 5.63 mGy), and is further confirmed by the corresponding GEE analysis (Table 4.2). It is also worthwhile to mention that an overall lower PB was observed at lower dose level, which indicated a counter-intuitively better accuracy with lower dose. This trend, also noted in a previous study[34], is a particularity resulted from the unstable performance of the segmentation software at low dose/high noise condition. We speculate that the software was more unstable when segmenting nodules with noisy boundaries and reported smaller nodule volumes. This decrease in reported nodule volume seemed to coincide with the overestimated volume, thus resulting in "better" accuracy at lower dose levels. As such, this particularity is not an indicator of more accurate volume estimation at lower dose, but rather an emphasis of the importance of sufficient but not excessive dose to yield robust segmentations for this particular software (>50% of current clinical dose level was indicated from Figure 4.6 (a), as statistical comparisons performed between successive dose levels showed insignificant differences between dose levels of 50% and above).

Figure 4.6 (b) shows the precision of the volume quantified with 2.5 mm slice thickness. Overall, the PRC drops quickly from 3% to 50% dose levels, and plateaus afterwards, showing a limited improvement in precision with increasing dose for all reconstruction algorithms. At 50-100% dose levels, all three reconstruction algorithms have similar PRC values. At dose levels less than 50%, both ASiR and MBIR have

smaller PRC values, i.e., better precision than FBP. Corresponding GEE analysis also shows a marginally significant improvement in precision with ASiR at 3% and 25% dose levels (Table 4.2).

Figure 4.6 (c-f) further shows the accuracy and precision of volume quantification with two thinner slice thicknesses: 1.25 and 2.5 mm. The accuracy of volume quantification with 1.25 and 0.625 mm slice thickness are shown in Figure 4.6 (c) and (e), respectively. Overall, the accuracy of the quantification improves by around 30% with thinner slice thicknesses. At 1.25 mm slice thickness, the difference between ASiR and FBP are greatly reduced as compared to 2.5 mm slice thickness, but still noticeable. MBIR remains comparable to FBP. GEE analysis shows that ASiR has significantly worse accuracy than FBP at most dose levels, and MBIR has accuracy similar to FBP except at the lowest dose levels (Table 4.2). At 0.625 mm slice thickness, the difference between the accuracy of ASiR and FBP are further reduced to be insignificant, as shown visually in Figure 4.6 (e) and via GEE analysis. The accuracy of MBIR, however, is greatly improves at higher dose levels, and is significantly better than the other two algorithms across most dose levels (Table 4.2). Figure 4.6 (d) and (f) shows the precision of volume quantification with 1.25 and 0.625 mm slice thickness, respectively. Overall, the precision of all three reconstructions at thinner slice thicknesses are slightly improved as compared to that of 2.5 mm slice thickness. IR's precision is still comparable to that of FBP at 1.25 and 0.625 mm slice thicknesses. The

only exception is the precision of MBIR at extremely low dose level with 0.625 mm slice, where the GEE analysis results show a marginal improvement with MBIR (Table 4.2).

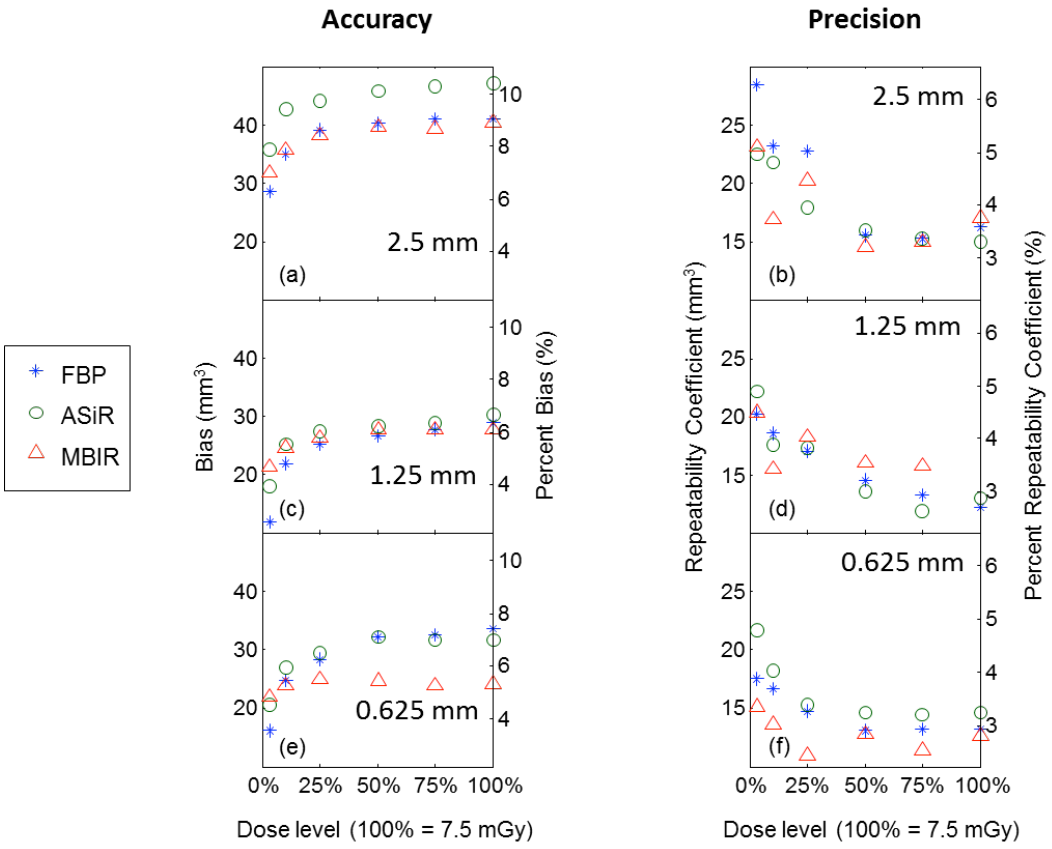

Figure 4.6: Accuracy and precision of 9.53 mm nodules segmented with Software A. (a,c,e) Accuracy at (a) 2.5 mm, (c) 1.25 mm, and (e) 0.625 mm slice thicknesses. (b,d,f) Precision at (b) 2.5 mm, (d) 1.25 mm, and (f) 0.625 mm slice thicknesses. The left y-axes show the absolute accuracy and precision in the unit of mm3, and the right y-axes show the relative accuracy and precision in the unit of %.

Table 4.2: Plotted GEE analysis results of the differences between FBP and IR in the volume quantification of 9.53 mm nodules with Software A (see Figure 4.6). Each solid circle represents a statistically significant difference with a p-value less than 0.05. Each open circle represents a marginally significant difference with a p-value between 0.05 and 0.1. All blank areas

indicate no statistically significant difference between the two reconstruction algorithms.

Accuracy, 9.53 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		3%	10%	25%	50%	75%	100%
2.5 mm	ASiR vs. FBP	●	●	●	●	●	●
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP	●	○	●	●		○
	MBIR vs. FBP	●	○				
0.625 mm	ASiR vs. FBP	○					
	MBIR vs. FBP			●	●	●	●
Precision, 9.53 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		3%	10%	25%	50%	75%	100%
2.5 mm	ASiR vs. FBP	○		○			
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP						
	MBIR vs. FBP						
0.625 mm	ASiR vs. FBP						
	MBIR vs. FBP		○				

Figure 4.7 summarizes the accuracy and precision of a smaller nodule – 4.76 mm in diameter. Compared to Figure 4.6, smaller nodules have overall worse accuracy and precision and noisier results due to the segmentation difficulty and the smaller number of nodules employed. Also compared to Figure 4.6, the counter-intuitively better

accuracy observed at low dose levels was less noticeable with smaller nodules, as smaller nodules have less nodule boundaries. Figure 4.7 (a) shows the accuracy of 4.76 mm nodule measured with 2.5 mm slice thickness. ASiR shows the worst accuracy, and MBIR shows the best. The corresponding GEE analysis results shown in Table 4.3 were not able to detect these observed differences to be statistically significant at 0.1 or lower level. This could be due to the low power resulted from small sample size (12 small nodules) or large uncertainty associated with the observed values. Figure 4.7 (b) shows the precision measured with 2.5 mm slice thickness, which suggests better precision with IR at lower dose levels, but better precision with FBP at higher dose levels. The following GEE analysis showed that only MBIR at 10% dose had significantly better precision than FBP (Table 4.3).

Figure 4.7 (c-f) further shows the accuracy and precision of 4.76 mm nodule with 1.25 and 0.625 mm slice thicknesses. Overall, the accuracy of 4.76 mm nodules improves 30-50% with thinner slice thicknesses, but the precision is not much affected. When comparing reconstruction algorithms, MBIR shows better accuracy than the other two algorithms at most dose levels, as shown in Figure 4.7 (c) and (e). ASiR's accuracy is worse than that of FBP at 1.25 mm, but comparable at 0.625 mm. The GEE analysis confirms some of the significance at dose levels ≤50%, as shown in Table 4.3. Figure 4.7 (b) and (d) shows the precision of the volume quantification with thinner slice thickness. At 1.25 mm slice thickness, MBIR shows worse but not significant precision than the

other two algorithms. At 0.625 mm slice thickness, no obvious trend is observed between the precision of IR and FBP. The GEE analysis, however, shows a marginally significant improvement in precision with IR at 20% and 50% dose (Table 4.3).

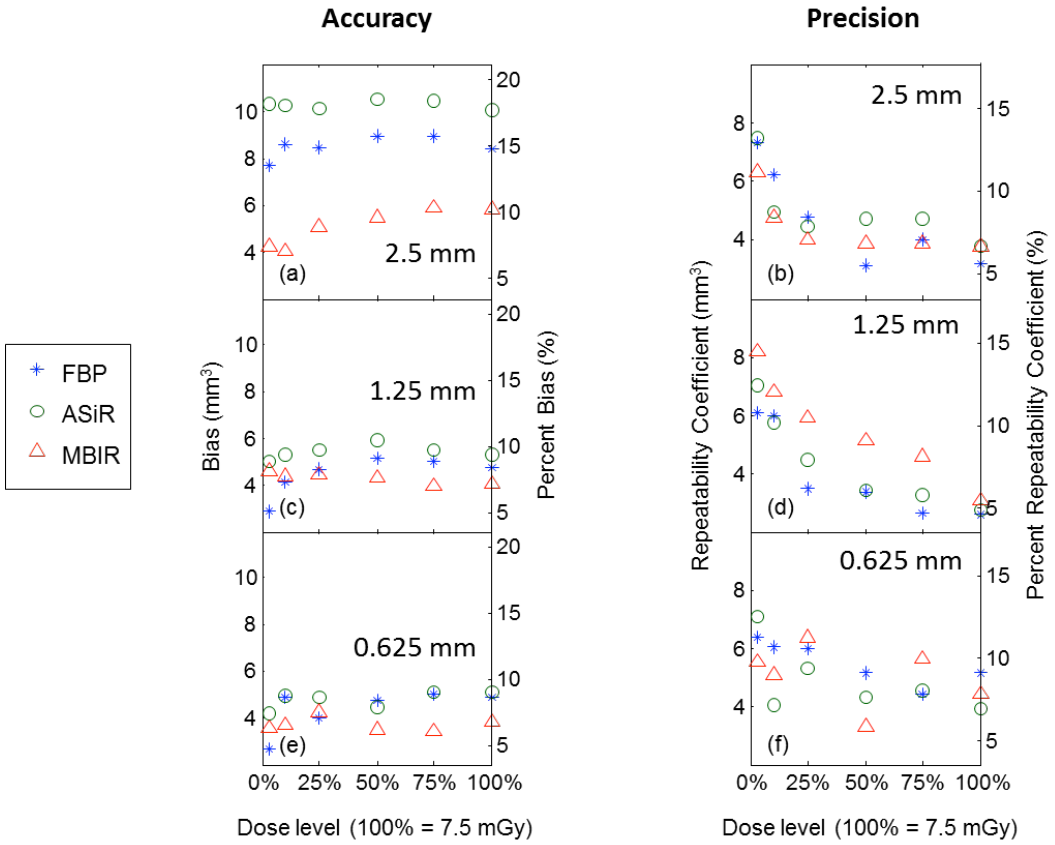

Figure 4.7: Accuracy and precision of 4.76 mm nodules segmented with Software A. (a,c,e) Accuracy at (a) 2.5 mm, (c) 1.25 mm, and (e) 0.625 mm slice thicknesses. (b,d,f) Precision at (b) 2.5 mm, (d) 1.25 mm, and (f) 0.625 mm slice thicknesses. The left y-axes show the absolute accuracy and precision in the unit of mm3, and the right y-axes show the relative accuracy and precision in the unit of %.

Table 4.3: Plotted GEE analysis results of the differences between FBP and IR in the volume quantification of 4.76 mm nodules with Software A (see Figure 4.7). Each solid circle represents a statistically significant difference with a p-value less than 0.05. Each open circle represents a marginally significant difference with a p-value between 0.05 and 0.1. All blank areas

indicate no significant difference between the two reconstruction algorithms.

Accuracy, 4.76 mm nodule

Slice thickness	Reconstruction algorithms	Dose level					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP						
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP	●	○	●	●		○
	MBIR vs. FBP	●	○				
0.625 mm	ASiR vs. FBP	○					
	MBIR vs. FBP			●	●	●	●

Precision, 4.76 mm nodule

Slice thickness	Reconstruction algorithms	Dose level					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP						
	MBIR vs. FBP		●				
1.25 mm	ASiR vs. FBP						
	MBIR vs. FBP						
0.625 mm	ASiR vs. FBP						
	MBIR vs. FBP		○				

4.3.3 Accuracy and precision of volume quantifications – Software B

The accuracy and precision of 9.53 mm nodule measured with Software B are shown in Figure 4.8 and Table 4.4. Software B generally followed the same trends as

observed with Software A: 1) The accuracy of MBIR was superior to that of FBP. 2) The precision of both ASiR and MBIR were mostly not significantly different from that of FBP. 3) Smaller nodule volumes were reported at lower dose levels. However, two deviations were noted: First, the accuracy of ASiR was equivalent or inferior to that of FBP with Software A, but superior to that of FBP with Software B. Secondly, the nodule volumes were generally overestimated with Software A, but underestimated with Software B (negative bias values in Figure 4.8). It is difficult to speculate the reason for these software dependencies. Nevertheless, they highlight a need to incorporate the "operator" in the characterization of volume quantification, and to implement some type of calibrations for software.

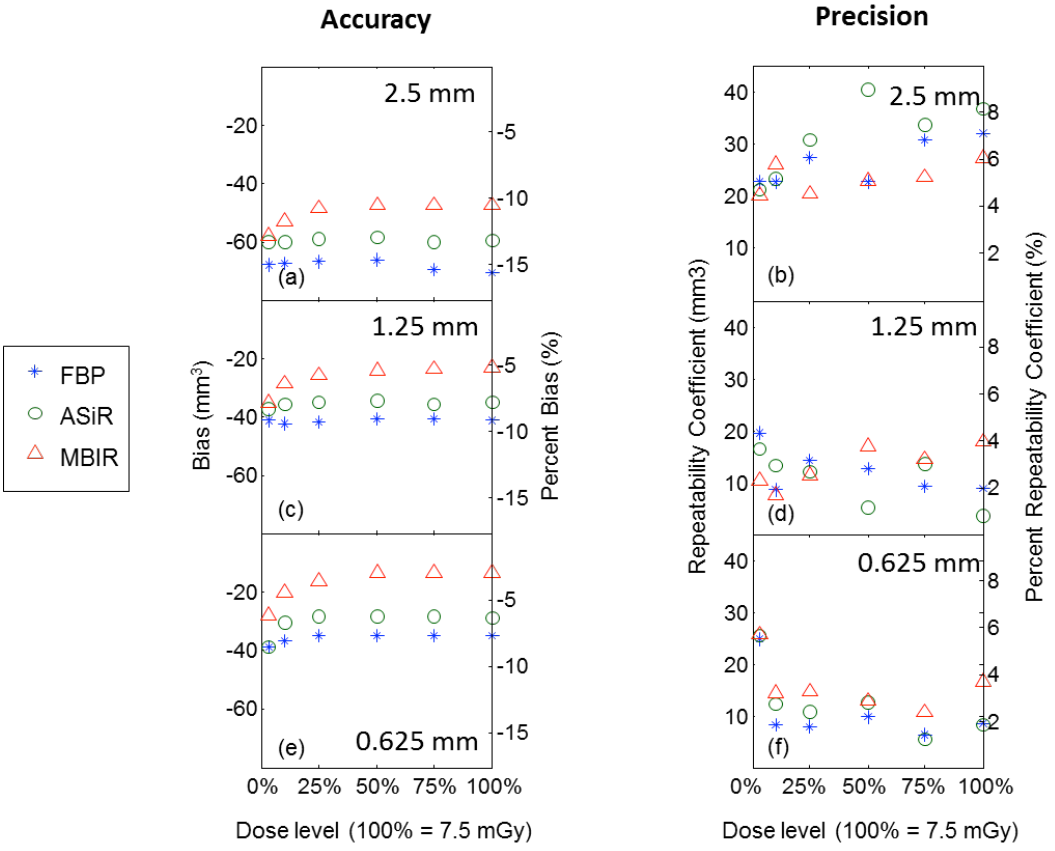

Figure 4.8: Accuracy and precision of 9.53 mm nodules segmented with Software B. (a,c,e) Accuracy of FBP and IR at (a) 2.5 mm, (c) 1.25 mm, and (e) 0.625 mm slice thicknesses. (b,d,f) Precision of FBP and IR at (b) 2.5 mm, (d) 1.25 mm, and (f) 0.625 mm slice thicknesses. The left y-axes show the absolute accuracy and precision in the unit of mm3, and the right y-axes show the relative accuracy and precision in the unit of %.

Table 4.4: Plotted GEE analysis results of the differences between FBP and IR in the volume quantification of 9.53 mm nodules with Software B (see Figure 4.8). Each solid circle represents a statistically significant difference with a p-value less than 0.05. Each open circle represents a marginally significant difference with a p-value between 0.05 and 0.1. All blank areas indicate no significant difference between the two reconstruction algorithms.

Accuracy, 9.53 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP	●	●	●	●	●	●
	MBIR vs. FBP	●	●	●	●	●	●
1.25 mm	ASiR vs. FBP	○	●	●	●	●	●
	MBIR vs. FBP	●	●	●	●	●	●
0.625 mm	ASiR vs. FBP		●	●	●	●	●
	MBIR vs. FBP	●	●	●	●	●	●
Precision, 9.53 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP			○			
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP						
	MBIR vs. FBP	○					○
0.625 mm	ASiR vs. FBP					●	
	MBIR vs. FBP		○				●

The accuracy and precision of 4.76 mm nodule measured with Software B generally followed the trends observed with Software A, as shown in Figure 4.9 and Table 4.5. The accuracy of both ASiR and MBIR were similar to that of FBP (a slightly higher PB with MBIR was observed at 1.25 mm slice thickness, but not statistically

significant). The precision of both ASiR and MBIR were not significantly different from that of FBP.

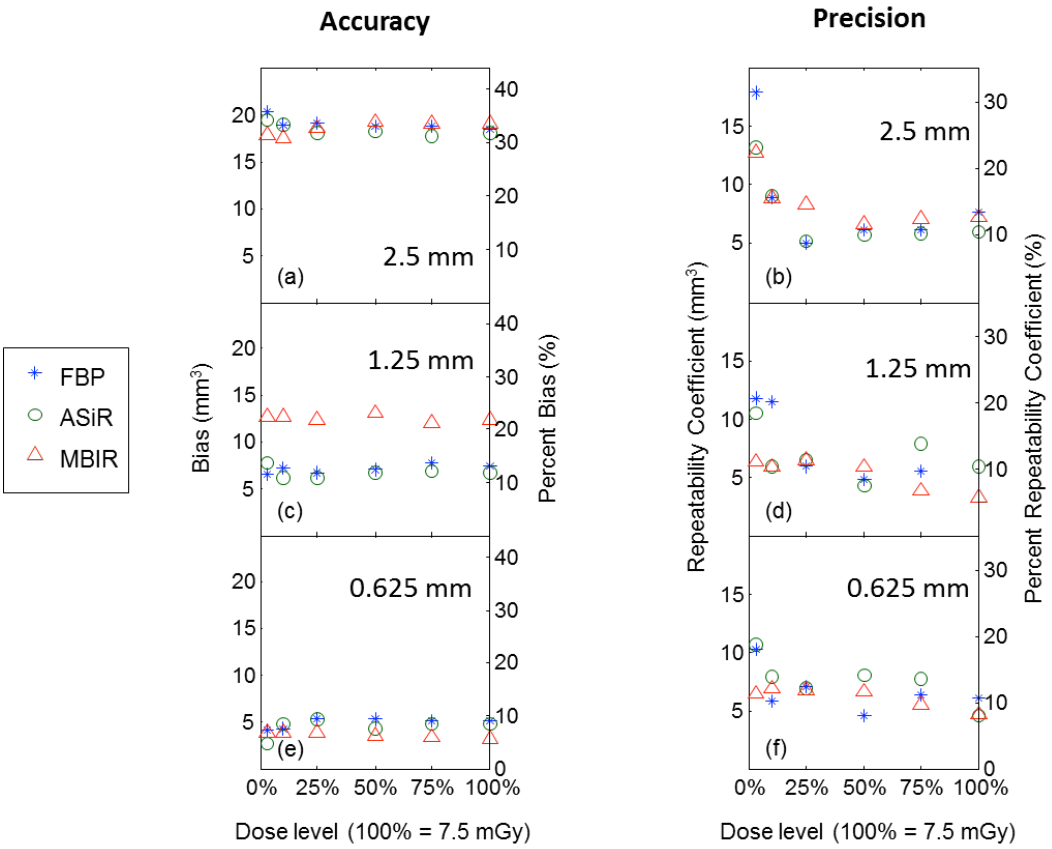

Figure 4.9: Accuracy and precision of 4.76 mm nodules segmented with Software B. (a,c,e) Accuracy of FBP and IR at (a) 2.5 mm, (c) 1.25 mm, and (e) 0.625 mm slice thicknesses. (b,d,f) Precision of FBP and IR at (b) 2.5 mm, (d) 1.25 mm, and (f) 0.625 mm slice thicknesses. The left y-axes show the absolute accuracy and precision in the unit of mm3, and the right y-axes show the relative accuracy and precision in the unit of %.

Table 4.5: Plotted GEE analysis results of the differences between FBP and IR in the volume quantification of 4.76 mm nodules with Software B (See Figure 4.9). Each solid circle represents a statistically significant difference with a p-value less than 0.05. Each open circle represents a marginally significant difference with a p-value between 0.05 and 0.1. All blank areas indicate no significant difference between the two reconstruction algorithms.

Accuracy, 4.76 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP						
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP						
	MBIR vs. FBP						
0.625 mm	ASiR vs. FBP						
	MBIR vs. FBP		○				

Precision, 4.76 mm nodule							
Slice thickness	**Reconstruction algorithms**	**Dose level**					
		0.03	0.1	0.25	0.5	0.75	1
2.5 mm	ASiR vs. FBP						
	MBIR vs. FBP						
1.25 mm	ASiR vs. FBP						
	MBIR vs. FBP						
0.625 mm	ASiR vs. FBP					○	
	MBIR vs. FBP						

4.4 Discussion

Assessing the nodule size change in terms of volume can potentially offer improved quantification as compared to traditional uni- and bi-dimensional methods. The usefulness of the volume quantification technique, however, depends on its accuracy and precision, which is a function of the protocol used. In this study, we investigated the impact of two iterative reconstruction algorithms (ASiR and MBIR) on the accuracy and precision of volume quantification across a range of dose levels, slice thicknesses, nodule sizes, and segmentation software.

This study found systematic difference between the accuracy of FBP and IR across two segmentation software: Software A yielded a consistently equivalent or inferior accuracy with ASiR as compared to FBP, and a consistently equivalent or superior accuracy with MBIR; Software B yielded a consistently equivalent or superior accuracy with ASiR, and a consistently equivalent or superior accuracy with MBIR for large nodules. While this reflects the performance of the three reconstruction algorithms in this particular study, it is worthwhile to point out that the segmentation software employed was designed and optimized for FBP images, not ASiR and MBIR images which are known to have different resolution and noise texture[24]. Therefore, it is possible that the difference between FBP and IR was due to the software being not compatible with IR image characteristics. Expanding current segmentation software to take IR

image characteristics into consideration will be beneficial. Furthermore, in case a follow-up scan is reconstructed with a different algorithm, a baseline calibration may aid to correct for the systematic offset, such that conclusions can be drawn on the growth/shrinkage of the nodule over time.

For both Software A and B, smaller nodule volumes were reported at lower dose/higher noise levels. This was counter-intuitive because image noise distributed symmetrically around zero, and was therefore expected to have a zero-sum impact on the volume segmentation. We speculate that this noise dependency may be a result of the morphological process employed in the segmentation algorithms, which preferably segments nodules into spherical shapes with smooth surfaces. When the noise magnitude reaches a certain level, the software's ability to determine a smooth nodule boundary may be impeded, resulting in a decreased volume estimate. This may also explain Software A's "better" accuracy at lower dose levels (Figure 4.6), as the decreased volume estimate cancels Software A's volume over-estimation. However, question remains in that why the noise-induced counter-intuitive trend was also observed in IR images, which, according to

Table **4.1**, had significantly reduced noise by 30-80% in the uniform region. The reason is that IR did not equally reduce the noise around nodule boundaries. As shown in the noise maps in Figure 4.10, IR algorithms (especially MBIR) greatly reduce noise in

the uniform region as compared to FBP, but barely reduce noise at the edges of lung structures, soft tissue, and bone. Quantitatively, at the 3% dose level, the noise in the uniform region of the noise maps were 73, 51, and 15 HU, for FBP, ASiR, and MBIR, respectively. Corresponding values at the edge region of the noise maps (calculated with an edge-detecting mask) were 63, 68, and 44 HU. With the edge of the nodule remained noisy with IR, the segmentation software reported smaller volumes at lower dose levels.

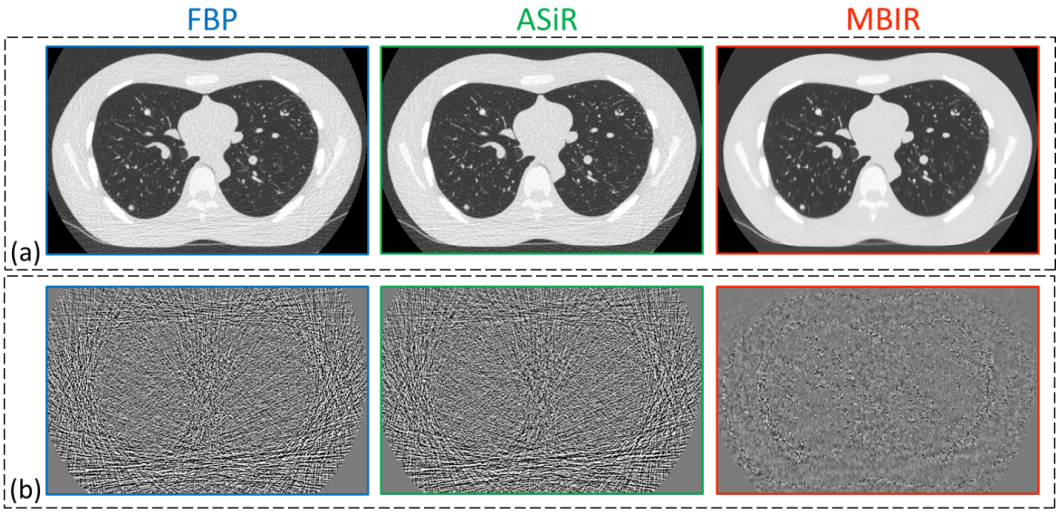

Figure 4.10: (a) The phantom image scanned at 3% dose level and reconstructed with 0.625 mm slice thickness. (b) The noise map corresponding to the phantom images, acquired by subtracting slices at the same axial location in repeated scans (the magnitude of the subtracted noise was corrected by a factor of square root of 2).

The precision of quantification was expected to improve with reduced image noise. In this study, despite ASiR and MBIR's notable noise reduction, similar precision

was found between FBP and IR. On the other hand, higher dose levels, which also offer reduced image noise, improved precision in most cases. The fact that IR's improved noise did not translate into improved precision may again be attributed to IR's enhanced noise at nodule edge (Figure 4.10). Noisier edges (regardless of reconstruction) would lead to degraded precision. Furthermore, the two IR employed in this study reduced noise at the cost of decreased resolution at extremely low dose levels[24], which also nullifies part of IR's noise reduction benefit.

As mentioned before, not all nodules were successfully segmented: at 25%-100% dose levels with MBIR reconstructions, two 4.76 mm nodules could not be segmented by Software A. Figure 4.11 illustrates one such nodule. The nodule was heavily cluttered with vessels. The segmentation of the nodule succeeded with FBP images, but failed with MBIR images at the same imaging condition. In the corresponding three-dimensional surface rendering of the nodule, MBIR images showed more vessels than FBP due to its enhanced edge definition and reduced noise. We therefore speculate that the segmentation failure of MBIR was a result of its increased vessel visibility, which somehow distracted the segmentation.

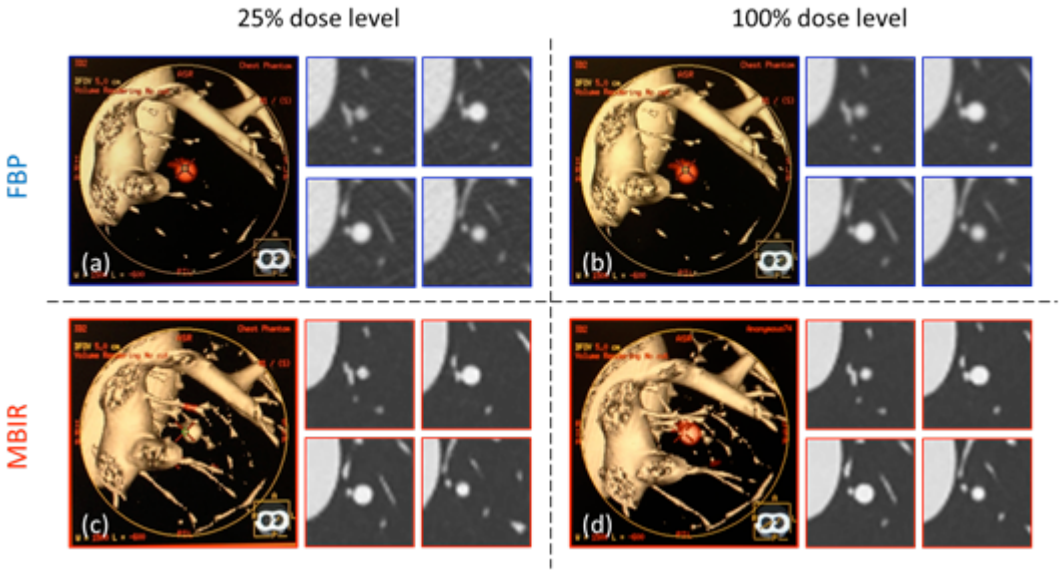

Figure 4.11: Three dimensional surface rendering and CT slices of the 4.76 mm nodule that (a,b) succeeded in the segmentation with FBP but (c,d) failed in the segmentation with MBIR.

Previously with FBP images, the accuracy of volume quantification was shown to be worse at 2.5 mm slice thickness than at 1.25 and 0.625 mm slice thicknesses[34]. In this study, at 2.5 mm slice thickness, similar accuracy deterioration was observed with FBP images. Furthermore, the deterioration was also observed with ASiR and MBIR images, indicating that a thinner slice thickness as an important factor for quantification accuracy regardless of the reconstruction algorithm.

As with most studies, our study had certain limitations. First, due to the capacity of the vesselled phantom, the nodules employed in this study were limited in terms of both type and number. Further tests will be beneficial to re-ascertain the impact of IR observed in this study with more nodules of different shapes and types. Tests with

increased number of nodules or repeated scans will also help compensate for the uncertainty caused by the nature of the vessel attachment, especially for small nodules. Second, it is very challenging to claim full generalizability from a limit set of image data with non-linear properties. However, even though this study does not fully examine every possible factor in volume quantification, it provides a methodology to examine quantitative tasks, and reveals trends and factors that have not been known before. Furthermore, some observed behaviors are generalizable (e.g., effect of slice thickness, dose, and behavior across segmentations). Therefore, we believe this study is meaningful and reasonably representative. Finally, both segmentation software employed in this study are commercial software, so that the quantitative results in this study can be reproduced by other scientists using the same software. However, current commercial software are mostly designed for FBP, thus might be sub-optimal for IR. Future works are warranted to investigate IR with open-source software that gives user more control of the segmentation process, and to develop generic segmentation software that are compatible with both FBP and IR images.

In summary, this study examined and compared the accuracy and precision of volume quantification with both FBP and IR algorithms. Both FBP and IR showed worse accuracy at 2.5 mm slice thickness. For large nodules, the accuracy was significantly different between ASiR and FBP for all slice thicknesses with both software, and significantly different between MBIR and FBP for 0.625 mm slice thickness with

Software A and for all slice thicknesses with Software B. For small nodules, the accuracy was more similar between FBP and iterative reconstruction, with the exception of ASIR vs. FBP at 1.25 mm with Software A and MBIR vs. FBP at 0.625 mm with Software A. The different accuracy between IR and FBP highlights the importance of extending the current segmentation software for the image characteristics of iterative reconstructions. In absence of software extension, a calibration process may help correct for the difference between the accuracy of FBP and IR, such that repeated scans with different reconstruction algorithms can be compared against each other. Few significant differences in precision were found between IR and FBP, possibly due to IR's diminished noise reduction at nodule edges and IR's degraded resolution at high noise conditions, but this does not rule out potential advantage of IR that might be evident in a study that uses a larger number of nodules or repeated scans.

5. Mathematical Prediction of Quantification Performance with IR*

5.1 Introduction

The high axial resolution images of multi-detector computed tomography (CT) have enabled the quantification of lung nodule volume, which is an important biomarker for cancer diagnosis and treatment response monitoring[33]. For example, the change of nodule volume over time, assessed from serial scans of the patient can be employed to ascertain treatment response and to differentiate benign and malignant nodules[78]. The usefulness of a volume quantification technique, however, depends on the precision of the quantification, which is the degree to which repeated quantifications of the same nodule under unchanged conditions yield the same outcome. To confidently assess nodule volume change, it is important to have a detailed knowledge of the quantification precision. Furthermore, since the precision might depend on acquisition parameters and nodule characteristics, the assessment needs to be performed across a wide range of conditions.

In recent years, a number of studies have been performed to assess the quantification precision experimentally[33, 34, 75, 79, 80]. The experimental method is effective, but extremely time consuming, as it involves multiple steps of case preparation, repeated scans, nodule segmentation, and statistical analysis. In addition, due to the

* This chapter is based on a manuscript submitted to *Medical Physics*.

complexity of CT acquisition and reconstruction parameters (dose, pitch, kernel, slice thickness, etc.), the experimental results are difficult to generalize.

To assess the quantification precision in an efficient and generalizable manner, a mathematical model named the estimability index (e') was developed by Richard and Samei[37]. The e' predicted the quantification precision by modeling the nodule and image characteristics in Fourier domain. It was shown to be effective in predicting the theoretical quantification precision obtained via a maximum likelihood estimator. However, in that implementation of e', for each nodule characteristics, the model needed to be first trained with pre-known precision values, therefore limiting the applicability of the model to a wide range of nodule characteristics. Furthermore, the model assumed ideal quantification software, which had *a priori* knowledge of the nodule's physical properties (size, contrast, shape, and edge profile), and made optimal use of the knowledge to segment the nodule.

Inspired by the prior work and motivated by its limitations, this study aimed to extend the e' model by replacing the nodule-specific training process with a more efficient, physically-based modeling process; the imperfection of the quantification software was modeled in terms of the discrepancy between the actual nodule and the segmentation software's expectation. Furthermore, the precision derived from the reformulated e' model was validated against empirical precision across 54 distinct acquisition protocols, 2 nodule sizes, and 2 segmentation software.

5.2 Methods and Materials

5.2.1 Estimability index (e')

A surrogate of volume quantification precision, the estimability index (e'), was developed by mathematically modeling the three factors influencing the quantification performance: the image quality (noise and resolution), the nodule characteristics, and the volume segmentation software.

The noise and resolution properties of the image were characterized in terms of noise power spectrum (NPS) and task transfer function (TTF), respectively. NPS is the square of the image noise (variance) as a function of spatial frequency, which describes both the magnitude and the texture of the noise. TTF is an extension of the modulation transfer function (MTF) to accommodate potential non-linearity of iterative reconstruction algorithms by describing the image resolution as a function of object contrast and image noise[24, 61]. In this study, both NPS and TTF were measured in three dimensions from a previously developed phantom, named the Mercury Phantom[38] (Figure 5.1). The three dimensional NPS was measured from the uniform region of the phantom; the in-plane TTF was measured from the rod inserts of various contrast levels using an edge technique; and the axial TTF was measured from the interfaces of various contrast levels between sections of the supplemental phantom, also using an edge technique. Both NPS and TTF measurements were described in detail in Chapter 2 and in previous studies [24, 38, 61], and have been validated for their accuracy and precision.

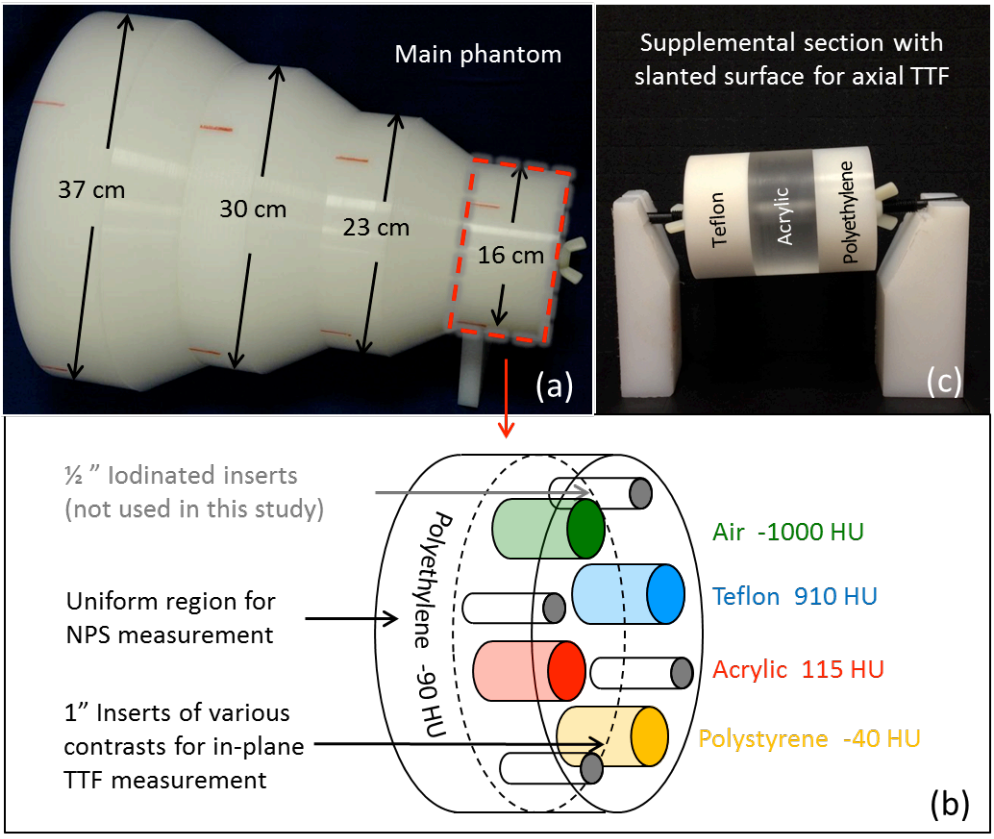

Figure 5.1: The three dimensional noise and resolution properties of the imaging system are measured from Mercury Phantom in terms of NPS and TTF, respectively. (a) The phantom is composed of four cylindrical sections with three tapered sections in between. (b) Each cylindrical section is divided into two subsections for the measurements of NPS and in-plane TTF. (c) A supplemental section with slanted surfaces provides the measurement of TTF along axial direction.

The second factor of quantification, the physical properties of the nodule, was mathematically modeled in terms of task function, W_{task}. The task function is the 3D Fourier transform of the nodule's edge profile, containing information such as the size, contrast, and edge profile of the nodule (Figure 5.2). The magnitude of the task function

is adjusted with a scale factor, such that the power of the function (the integral of the task function in frequency domain) equals the power of the nodule (the integral of the nodule in spatial domain).

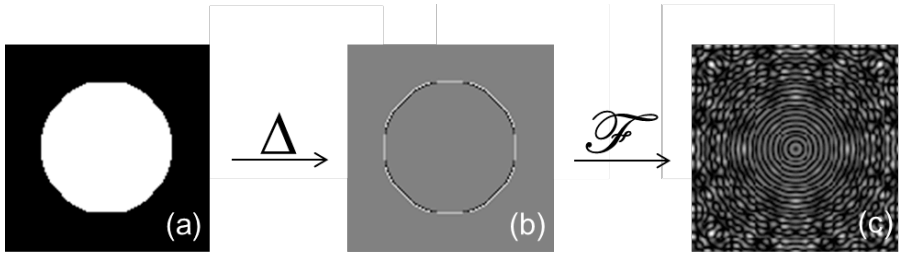

Figure 5.2: (a) A nodule is mathematically modeled. (b) The edge profile of the nodule is detected using a discrete Laplace operator. (c) The task function of the nodule is calculated as the Fourier transform of the edge profile. The nodule and its task function are three-dimensional, but plotted in two-dimensions for display purpose.

The last component of the quantification precision, the nodule segmentation process, was modeled as a cross-correlation between the nodule and a template, described in Fourier domain as a template function, W_{temp}. If the template matches the nodule, the segmentation is optimized; if not, the segmentation is biased towards the template. The morphological processes employed by most commercial segmentation software favor spherical or lobular nodules[81]. As a result, for most spherical nodules, the template matches the nodule well. However, speculated nodules are penalized as they do not fit the spherical assumption of most segmentation algorithms.

Finally, inspired by the non-prewhitening matched filter model observer employed for detection tasks (detectability index)[28], the e' was formulated as

$$\frac{1}{e'^2} = \frac{\iiint NPS \cdot TTF^2 \cdot |W_{task}| \cdot |W_{temp}| \, dudvdw}{\left(\iiint TTF^2 \cdot |W_{task}| \cdot |W_{temp}| \, dudvdw\right)^2}, \quad (5.1)$$

where u, v, and w are orthogonal spatial frequencies, and C and N are the contrast and noise levels at which the nodule volumes are quantified. e' modeled the quantification (segmentation) process as a cross correlation between the nodule and the template, which is equivalent to the product between the template function and the task function in Fourier domain. The numerator represents the fluctuation of the segmentation due to image noise and the mismatch between the nodule and the template, while the denominator normalizes it by the strength of the cross-correlation, i.e., the similarity between the nodule and the template. Note that higher e' value represents larger fluctuation and therefore worse quantification precision. Higher noise (larger NPS), poorer resolution (lower TTF), or worse prior knowledge of nodule (mismatch between the template function and the task function) result in larger e'.

5.2.2 Experimental measurements of quantification precision

To verify our e' model, the gold standard of quantification precision was experimentally acquired in terms of percent repeatability coefficient (PRC), as described in detail in Chapter 4 and in a previous study[23]. PRC is the expected percentage difference between any two repeated quantifications of the same nodule, for 95% of cases[23, 76]. Smaller PRC therefore represents better quantification precision.

To calculate PRC, synthetic nodules (acrylic; 80 HU at 120 kVp) of 9.5 and 4.8 mm in diameters were embedded in an anthropomorphic chest phantom (LUNGMAN, KYOTO KAGAKU, Kyoto, Japan), attaching to lung vessel structures and pleura. The phantom was scanned repeatedly on a 64 slice CT scanner (Discovery CT750 HD, GE Healthcare, Waukesha, WI) with 54 distinct protocols, including 6 dose levels (100, 75, 50, 25, 10 and 3%, with 100% dose corresponding to a $CTDI_{vol}$ of 7.5 mGy), 3 slice thicknesses (0.625, 1.25, and 2.5 mm), and 3 reconstruction algorithms [Filtered Back-projection (FBP), Adaptive statistical Iterative Reconstruction (ASiR), and Model-based Iterative Reconstruction (MBIR)]. Each protocol was repeated 5 times for precision calculation. Nodule volumes were quantified from the phantom images using two semi-automatic clinical segmentation software (Software A: LungVCAR, GE Healthcare, Waukesha, WI; and Software B: iNtuition, TeraRecon, Foster City, CA). 108 PRC values were calculated from the quantified volumes, corresponding to the 54 acquisition protocols and the 2 nodule sizes.

5.2.3 Relating e' to experimental precision

e' was then calculated for the same acquisition protocols and nodule sizes employed in the PRC calculation. First, a library of TTF and NPS was derived from the Mercury Phantom measurements, characterizing the resolution and noise properties of the operating space under a range of dose levels, reconstruction algorithms, and slice thicknesses. Specific TTF and NPS relevant to the measurement condition of PRC were

then interpolated from the library according to the nodule contrast, the noise magnitude, the reconstruction algorithm, and the slice thickness of the chest phantom images. Because of IR's different noise magnitudes in textured region and in uniform region[23, 58], the noise magnitude used for interpolations was measured from the lung region. In that process, lung vessels were removed prior to the noise measurements by subtracting repeated scans and adjust the pixel intensity by a factor of $\sqrt{2}$. In total, 54 pairs of TTF and NPS were interpolated from the library to represent the resolution and noise properties of the 54 protocols used in PRC calculations. The task functions were modeled for the 2 types of nodules used for PRC calculations. The template functions were modeled as identical to the task functions since the nodules were perfectly spherical. Finally, e' was calculated across all protocols and nodule sizes, and related to PRC to see how it may predict quantification precision. Relationships were established between e' and PRC, which were further employed to develop a process for e'-based PRC prediction.

5.3 Results

5.3.1 Reconstructed images

As a visual example of how acquisition parameters impact the image appearance and further quantification precision, images of the anthropomorphic chest phantom reconstructed with three algorithms are shown in Figure 5.3(a-c). The same images with noise only (structure removed by subtracting repeated scans) are shown in Figure 5.3(d-

f). Both ASiR and MBIR showed noise reduction as compared to FBP. MBIR had more noise reduction in the relatively uniform areas of the images (e.g., soft tissue) than around lung vessels and nodules.

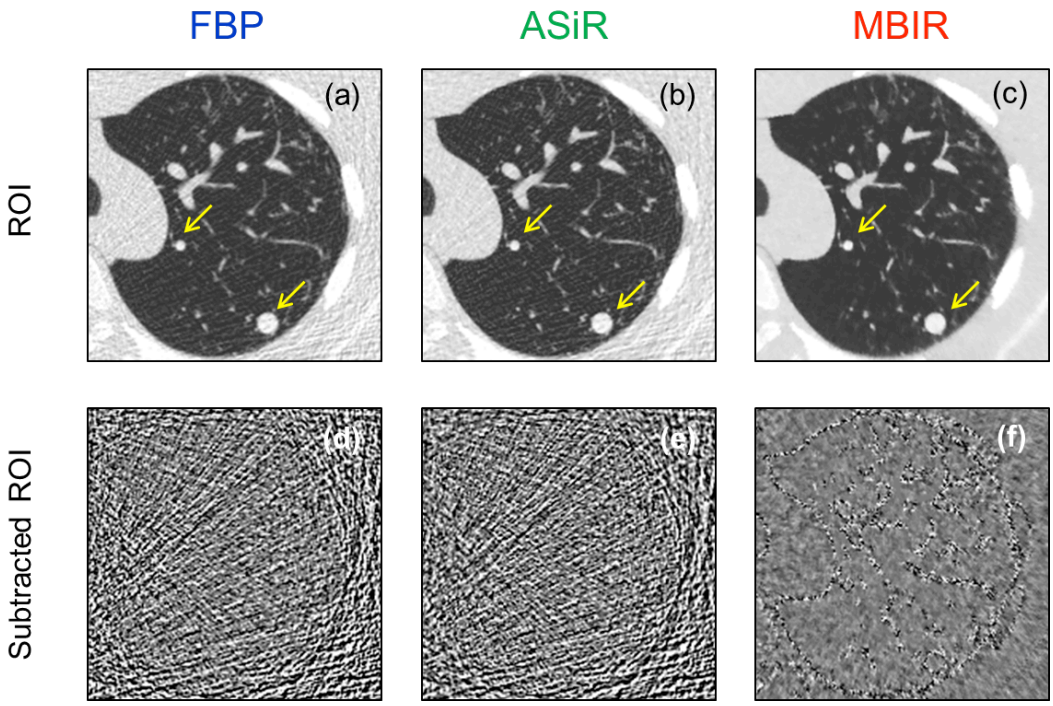

Figure 5.3: (a-c) Regions of interest showing the anthropomorphic chest phantom (used for PRC calculation) reconstructed at 0.625 mm slice thickness and 3% dose with three algorithms: FBP, ASiR, and MBIR. The nodules being quantified are highlighted with arrows. (d-f) Subtracted regions of interest showing the noise only.

5.3.2 TTF and NPS

Figure 5.4 shows the TTF and NPS measured at various dose levels for all three reconstruction algorithms, with the slice thickness fixed at 1.25 mm. Although three-dimensional TTF and NPS were used for e' calculation, Figure 5.4 only shows the in-

plane radially averaged results for visualization and comparison purposes. Compared to FBP, the two IR algorithms demonstrate enhanced TTF, i.e., better resolution, but also a noise dependency, with higher dose (lower image noise) levels yielding better resolution. The two IR algorithms showed NPS curves with smaller magnitude but lower peak frequency, reflecting IR's lower frequency noise feature, i.e., "waxier" appearance (as seen in Figure 5.3).

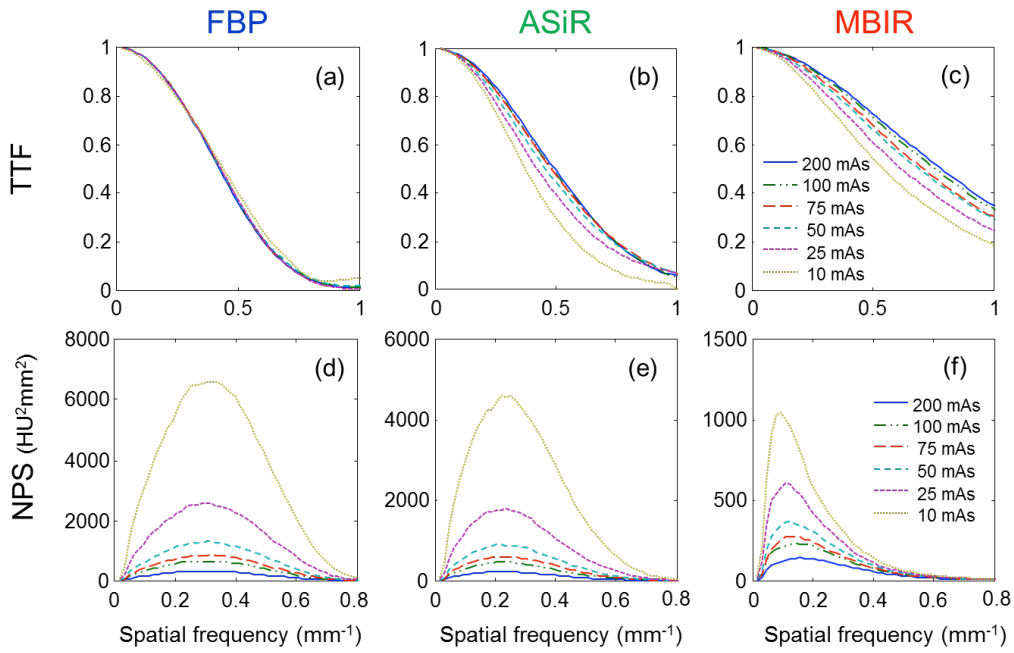

Figure 5.4: (a,b,c) The TTF of three reconstruction algorithms (FBP, ASiR, and MBIR) at various dose levels. The contrast level for TTF measurement is fixed at 1000 HU. (d,e,f) The NPS of the three reconstruction algorithms at various dose levels.

5.3.3 e' results and validation

Figure 5.5 shows the relationship between e' and PRC for 54 protocols, 2 nodule sizes, and 2 segmentation software. The relationships between e' and PRC exhibiting variability primarily due to the nature of the PRC quantification[23]. Nevertheless, they follow general non-linear relationships, fitted to the form of $PRC = a\ln\left(b \cdot e^{1^c} + 1\right)$, In general, the relationships are independent of nodule size, imaging dose, reconstruction algorithm, and slice thickness, but depend on the segmentation software, with Software A showing a stronger correlation than software B.

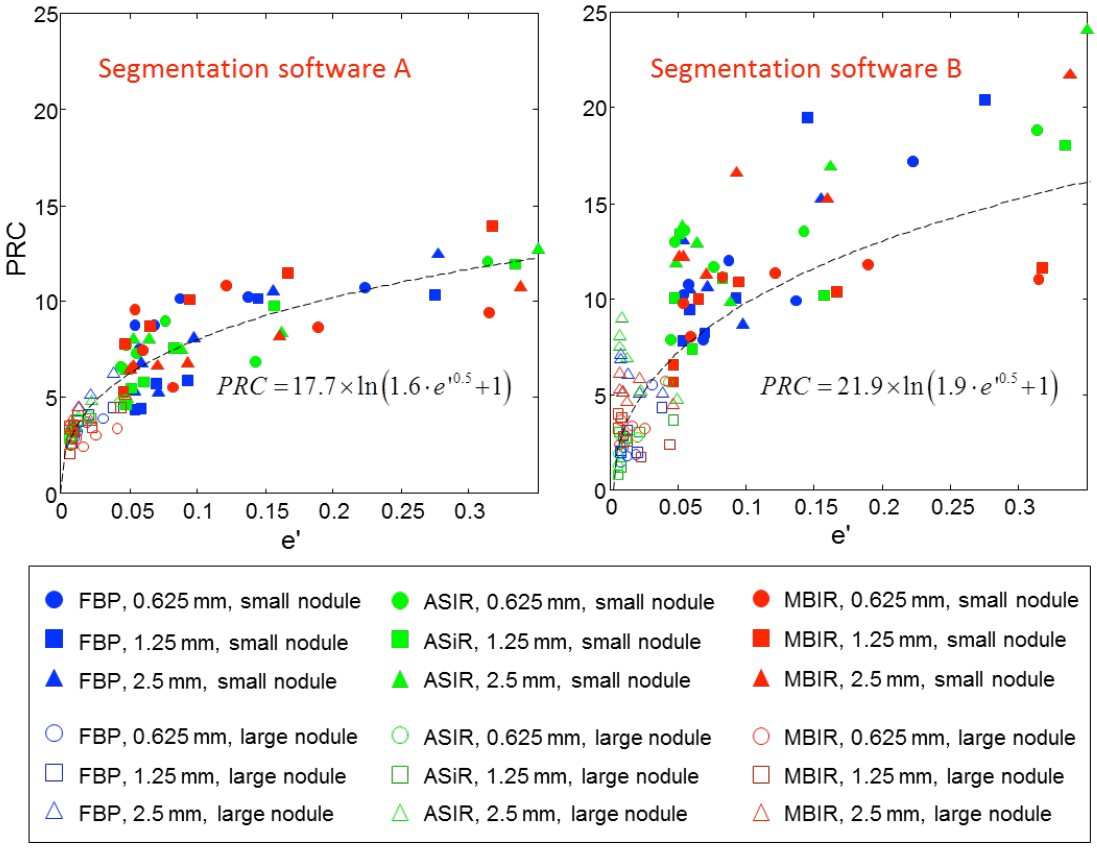

Figure 5.5: PRC verses 1/e' across 3 reconstruction algorithms, 3 slice thicknesses, 6 dose levels, 2 nodule sizes, and 2 segmentation software.

5.3.4 e' as a compliance tool

With the relationships established between e' and PRC, a process for e'-based PRC assessment is summarized in Table 5.1. The process is based on two steps. In Step 1, the image quality phantom is scanned several times to establish a library of TTF and NPS characterizing the resolution and noise properties of the operating space. Step 2 models the task function and the template function, and combines them with the library to calculate the PRC.

Table 5.1: Process for e'-based assessment of PRC.

Step	Element	Task
Step 1	**A**	Image quality phantom that contains - edges of various attenuations along various directions for three-dimensional *TTF* measurements - uniform volume for three-dimensional *NPS* measurements Scan the phantom at a range of dose levels and reconstruct the image with multiple slice thicknesses and reconstruction algorithms to compute a library of *TTF* and *NPS* that characterizes the operating space
Step 2	**B**	Model W_{task} and W_{temp} according to the nodule size, shape, and contrast
	C	Interpolate a 3D *TTF* from the library built in Step 1 with respect to nodule contrast and image noise (not necessary if the reconstruction algorithm is linear) Interpolate a 3D *NPS* from the library with respect to image noise
	D	Calculate e' according to *TTF*, *NPS*, W_{task} and W_{temp}
	E	Relate *e'* to *PRC* using the relationships established in Figure 5.5

The e'-based PRC assessment saved considerable time as compared to experimental-based PRC assessment. The experimental-based assessment in this study involved 30 scans (6 dose levels x 5 repeats) and 9 reconstructions per scan (3 reconstruction algorithms x 3 slice thicknesses). Furthermore, each dataset contained 35 nodules (21 9.5-mm nodule and 14 4.8-mm nodules) to be semi-manually segmented with 2 software, yielding a total of 18900 segmentations. Given that each segmentation took about 1 minute, the whole process toke more than 300 hours. On the contrary, the calculation of e' involved only 6 scans (6 dose levels) and 9 reconstructions per scan (3 reconstruction algorithms x 3 slice thicknesses) to establish a library of NPS and TTF, and mathematically calculated e' for all nodule types and acquisition protocols. The total

calculation time of e' calculation was around 3 hours (2 hour of data collection and 1 hour of data analysis).

5.4 Discussion

To quantify the development of lung nodule volumes with confidence, the precision of CT volume quantification needs to be assessed for a wide range of image acquisition and reconstruction parameters and nodule characteristics. The traditional experimental assessment of precision is time consuming. As a result, this study developed a mathematical model to predict quantification precision in terms of an estimability index (e'). Results showed a strong correlation between e' and empirical precision, indicating e' as an effective predictor of quantification precision.

As mentioned in Introduction, this e' model was an extension of a prior e' methodology[37]. One limitation of that work was that it relied on pre-known precision values to train the task function for each individual nodule type. In this study, by introducing the task function based on the principle of volumetry (edge detection), not only the computational burden of training process was eliminated, but also the possible bias introduced by the training data was avoided. Another limitation of the prior work was that it did not explicitly model the segmentation software as a influencing factor of precision performance. In this study, the imperfection of clinical segmentation software was modeled in two regards: first, a non-prewhitening matched filter model instead of a prewhitening model was used to capture the impact of noise correlation on

quantification precision[28]; second, a template function was introduced to account for the possible mismatch between the nodule and the software's expectation.

In clinical practices, methods other than volume quantification are also employed to assess the size of lung nodule, such as the bi-dimensional quantification suggested by World Health Organization (WHO)[71] and the uni-dimensional quantification suggested by Response Evaluation Criteria in Solid Tumors (RECIST) working group[70]. This study chose to focus on volume quantification because it has been shown to be more accurate and precise compared to the bi- and uni-dimensional quantifications, especially for complex lesion shapes[82]. To reflect the three-dimensional nature of volume quantification, the e' was modeled in three dimensions. However, similar e' models in one and two dimensions may be derived to evaluate the performance of quantifications using WHO and RECIST methods, if needed.

Two facts indicated imperfect modeling of the segmentation software in e': 1) the relationships established between e' and PRC were generally independent of image acquisition parameters and nodule characteristics, but depend on volume segmentation software; 2) the correlation between e' and PRC was stronger for Software A than for Software B. One possible solution is to replace the current software-generic template function with a software-specific template function, which closely models the segmentation algorithm of each software and its impact on quantification precision. Another possible solution is to include an "internal noise" term, which models the

inconsistency of the software during repeated segmentations of identical dataset, and counts it towards precision deterioration. These remain as possible future extensions of the e' methodology.

It should be pointed out that even though the correlations between e' and PRC were weaker for Software B than for Software A, it did not necessarily indicate that e' is worse at predicting the precision of quantifications performed with Software B. This is because PRC, the gold standard for quantification precision, even estimated with a challengingly large number of nodules (21 9.5-mm nodules and 14 4.8-mm nodules), had uncertainty due to the limited sample size[23]. Further studies with a larger sample size are warranted to confirm the PRC.

In conclusion, a mathematical model based on an estimability index was developed to assess the quantification precision efficiently. The model was shown effective across a range of image acquisition protocols, nodule sizes, and segmentation software, and can be employed to optimize clinical protocols for quantification tasks.

6. Conclusions, Implications, and Future Directions

6.1 Summary and Conclusions

This dissertation developed a framework for task-specific evaluations of IR performance. Two types of tasks were investigated: detection tasks and volume quantification tasks. For each type of task, a mathematical model was developed to evaluate IR effectively and efficiently.

As a basis for performance evaluation, an experimental methodology was first developed in Chapter 2 to characterize the noise and resolution properties in a manner that was relevant to IR and 3D volumetric imaging. Results showed that this method was efficient and meaningful in characterizing the system performance accounting for the non-linearity of IR at multiple contrast and noise levels. It was also shown that when certain criteria were met, the measurement error could be controlled to be less than 10% to allow challenging measuring conditions with low object contrast and high image noise.

Chapter 3 incorporated the noise and resolution characterizations developed in Chapter 2 into the d' calculations, and evaluated the performance of IR and FBP for detection tasks. Results showed that compared to FBP, IR required less dose to achieve a threshold performance level (AUC), therefore potentially reducing the required dose. The dose saving potential of IR was not constant, but dependent on the task properties, with subtle tasks (small size and low contrast) enabling more dose saving than

conspicuous tasks. Results also showed that at a fixed dose level, IR could detect more subtle tasks than FBP, demonstrating the overall superior performance of IR for detection tasks.

Chapter 4 experimentally compared the volume quantification performance of IR to FBP in terms of accuracy and precision. Results showed that across a wide range of dose and slice thickness, IR led to accuracy significantly different from that of FBP, highlighting the importance of calibrating or expanding current segmentation software to incorporate the image characteristics of IR. Results also showed that despite IR's great noise reduction in uniform regions, IR in general had quantification precision similar to that of FBP, possibly due to IR's diminished noise reduction at edges (such as nodule boundaries) and IR's loss of resolution at low dose levels.

Lastly, Chapter 5 extended the e' model from a prior study[37], and employed it to mathematically predict the quantification performance of IR and FBP. Results showed that the extended e' model was improved in three respects, including the task modeling, the segmentation software modeling, and the characterizations of noise and resolution properties. Results also showed that the extended e' model correlated with experimental precision (Chapter 4) across a range of image acquisition protocols, nodule sizes, and segmentation software. In addition, compared to experimental assessments of quantification performance, e' was significantly reduced in computational time, such

that it can be easily employed in clinical studies to verify quantitative compliance and to optimize clinical protocols for CT volumetry.

6.2 Clinical Implications

The research in this dissertation has two important clinical implications. First, because d' values reflect the percent of detection accuracy (AUC) and e' values reflect the quantification precision, this work provides a framework for evaluating IR with diagnostic accuracy as the endpoint. Second, because the calculations of d' and e' models are much more efficient compared to conventional observer studies, the clinical protocols with IR can be optimized in a timely fashion, and the compliance of clinical performance can be examined routinely.

6.3 Future Directions

In this dissertation the noise and resolution properties were measured from a phantom with uniform backgrounds. However, it is of interest to also assess the image noise and resolution properties in presence of background variability, because the textured backgrounds (such as lung vessels) and irregular objects (such as speculated lesions) might affect local noise and resolution properties of IR images. Therefore, phantoms with textured background and irregular shaped inserts are warranted in future studies to fully investigate the non-linear noise and resolution properties of IR.

Because the noise and resolution properties employed in d' calculations were measured from uniform backgrounds, the corresponding d' values represented the

detection performance of tasks in relatively uniform background, such as hepatic lesions and kidney stones. Furthermore, the d' model employed in this dissertation work did not accounted for the impact of anatomical noise on the performance of human observer. Therefore, to evaluate IR for tasks in texture background, future efforts are needed to measure noise and resolution properties from textured background and to incorporate anatomical noise into the d' model.

In the experimental evaluation of IR's quantitative performance, the type and number of synthetic nodules were limited (21x 9.5-mm acrylic nodules and 14x 4.8-mm acrylic nodules) by the capacity of the anthropomorphic phantom and the number of segmentations practicable in one study (35 nodules x 54 protocols x 5 repeats x 2 segmentation software = 18900 segmentations). Further studies will be beneficial to re-ascertain the observed impact of IR with nodules of variable density, shape, and size. Tests with increased number of nodules of each type will also help compensate for the inter-nodule variability caused by the nature of the vessel attachment, especially for small nodules.

Although e' in general agreed well with experimental precision, the relationships between the two were still dependent on volume segmentation software. In future extensions of the e' methodology, the software-generic template function may be replaced with a software-specific template function to closely model the segmentation algorithm of each software and its impact on quantification precision. An iterative

technique per Richard and Samei[37] may be used to empirically determine the template function for a segmentation algorithm, or a priori knowledge form the manufacturer can be incorporated into the model. An "internal noise" term may also be introduced to model the inconsistency of the segmentation software during repeated segmentations of identical dataset.

References

1 J. Nuyts, B. De Man, J.A. Fessler, W. Zbijewski, F.J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," Phys Med Biol **58**, R63-96 (2013).

2 J.A. Fessler, "Penalized Weighted Least-Squares Image-Reconstruction for Positron Emission Tomography," Ieee T Med Imaging **13**, 290-300 (1994).

3 A. Ziegler, T. Nielsen, M. Grass, "Iterative reconstruction of a region of interest for transmission tomography," Med Phys **35**, 1317-1327 (2008).

4 D.L. Snyder, C.W. Helstrom, A.D. Lanterman, M. Faisal, R.L. White, "Compensation for Readout Noise in Ccd Images," Journal of the Optical Society of America a-Optics Image Science and Vision **12**, 272-283 (1995).

5 G.H. Olivera, D.M. Shepard, P.J. Reckwerdt, K. Ruchala, J. Zachman, E.E. Fitchard, T.R. Mackie, "Maximum likelihood as a common computational framework in tomotherapy," Physics in Medicine and Biology **43**, 3277-3294 (1998).

6 J.B. Thibault, K.D. Sauer, C.A. Bouman, J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," Medical Physics **34**, 4526-4544 (2007).

7 K. Sauer, C. Bouman, "A Local Update Strategy for Iterative Reconstruction from Projections (Vol 41, Pg 534, 1993)," Ieee T Signal Proces **41**, 2022-2022 (1993).

8 T.M. Benson, B.K.B. De Man, L. Fu, J.B. Thibault, "Block-Based Iterative Coordinate Descent," 2010 Ieee Nuclear Science Symposium Conference Record (Nss/Mic), 2856-2859 (2010).

9 Z. Yu, J.B. Thibault, C.A. Bouman, K.D. Sauer, J.A. Hsieh, "Fast Model-Based X-Ray CT Reconstruction Using Spatially Nonhomogeneous ICD Optimization," Ieee T Image Process **20**, 161-175 (2011).

10 M. Willemink, T. Leiner, P. Jong, L. Heer, R.J. Nievelstein, A.R. Schilham, R.J. Budde, "Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality," European Radiology **23**, 1632-1642 (2013).

11 M. Willemink, P. Jong, T. Leiner, L. Heer, R.J. Nievelstein, R.J. Budde, A.R. Schilham, "Iterative reconstruction techniques for computed tomography Part 1: Technical principles," European Radiology **23**, 1623-1631 (2013).

12 R.C. Nelson, S. Feuerlein, D.T. Boll, "New iterative reconstruction techniques for cardiovascular computed tomography: how do they work, and what are the advantages and disadvantages?," J Cardiovasc Comput Tomogr **5**, 286-292 (2011).

13 W. Bogdanich, "Radiation Boom. Available via http://topics.nytimes.com/top/news/us/series/radiation_boom/index.html," in *Radiation Boom* (The New York Times, 2011).

14 K. Grant, T. Flohr, "Iterative Reconstruction in Image Space (IRIS). Siemens Medical Solutions Whitepaper. Available via http://www.medical.siemens.com/siemens/en_US/gg_ct_FBAs/files/Case_Studies/IRIS_Whitepaper.pdf," (2010).

15 K. Grant, R. Raupach, "SAFIRE: Sinogram Affirmed Iterative Reconstruction. Siemens Medical Solutions Whitepaper. Available via http://usa.healthcare.siemens.com/siemens_hwem-hwem_ssxa_websites-context-root/wcm/idc/siemens_hwem-hwem_ssxa_websites-context-root/wcm/idc/groups/public/@us/@imaging/@ct/documents/download/mdaw/ndq2/~edisp/safire-00308312.pdf," (2012).

16 J.-B. Thibault, "Veo: CT Model-Based Iterative Reconstruction. GE Healthcare Whitepaper. Available via http://www.gehealthcare.com/dose/pdfs/Veo_white_paper.pdf," (2010).

17 B. Irwan, S. Nakanishi, A. Blum, "AIDR 3D - Reduces Dose and Simultaneously Improves Image Quality. Toshiba Medical Systems Whitepaper. Available via http://www.toshiba-medical.eu/upload/TMSE_CT/White Papers/White Papers/Toshiba_White paper CT_nov11.pdf?epslanguage=en," (2012).

18 A. Scibelli, "iDose4 iterative reconstruction technique. Philips Healthcare Whitepaper. Available via https://http://www.healthcare.philips.com/pwc_hc/main/shared/Assets/Documents/ct/idose_white_paper_452296267841.pdf," (2011).

19 M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu, K. Ohtomo, "Model-based iterative reconstruction technique for radiation dose

reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique," European Radiology **22**, 1613-1623 (2012).

20 B. Chen, D. Marin, S. Richard, D. Husarik, R. Nelson, E. Samei, "Precision of Iodine Quantification in Hepatic CT: Effects of Iterative Reconstruction With Various Imaging Parameters," Am J Roentgenol **200**, W475-W482 (2013).

21 S. Singh, M.K. Kalra, J. Hsieh, P.E. Licato, S. Do, H.H. Pien, M.A. Blake, "Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques," Radiology **257**, 373-383 (2010).

22 A. Moscariello, R.A.P. Takx, U.J. Schoepf, M. Renker, P.L. Zwerner, T.X. O'Brien, T. Allmendinger, S. Vogt, B. Schmidt, G. Savino, C. Fink, L. Bonomo, T. Henzler, "Coronary CT angiography: image quality, diagnostic accuracy, and potential for radiation dose reduction using a novel iterative image reconstruction technique—comparison with traditional filtered back projection," European Radiology2011).

23 B. Chen, H. Barnhart, S. Richard, M. Robins, J. Colsher, E. Samei, "Volumetric quantification of lung nodules in CT with iterative reconstruction (ASiR and MBIR)," Medical Physics **40**, 111902 (2013).

24 S. Richard, D.B. Husarik, G. Yadava, S.N. Murphy, E. Samei, "Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms," Medical Physics **39**, 4115-4122 (2012).

25 B. Chen, S. Richard, O. Christianson, X. Zhou, E. Samei, "CT performance as a variable function of resolution, noise, and task property for iterative reconstructions," in *SPIE Medical Imaging*, *Vol. 8313* (SPIE, San Diego, CA, 2012), pp. 83131K.

26 J.-B. Thibault, K.D. Sauer, C.A. Bouman, J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," Medical Physics **34**, 4526-4544 (2007).

27 ICRU, "Medical imaging - the assessment of image quality," *Vol. 54* (1996).

28 A.E. Burgess, "Visual Perception Studies and Observer Models in Medical Imaging," Seminars in Nuclear Medicine **41**, 419-436 (2011).

29 G.J. Gang, D.J. Tward, J. Lee, J.H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," Medical Physics **37**, 1948 (2010).

30 S. Richard, J.H. Siewerdsen, "Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images," Medical Physics **35**, 5043-5053 (2008).

31 G.J. Gang, J. Lee, J.W. Stayman, D.J. Tward, W. Zbijewski, J.L. Prince, J.H. Siewerdsen, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance," Medical Physics **38**, 1754 (2011).

32 A.E. Burgess, B. Colborne, "Visual signal detection. IV. Observer inconsistency," J Opt Soc Am A **5**, 617-627 (1988).

33 M.A. Gavrielides, L.M. Kinnard, K.J. Myers, N. Petrick, "Noncalcified lung nodules: volumetric assessment with thoracic CT," Radiology **251**, 26-37 (2009).

34 B. Chen, H. Barnhart, S. Richard, J. Colsher, M. Amurao, E. Samei, "Quantitative CT: technique dependence of volume estimation on pulmonary nodules," Phys Med Biol **57**, 1335-1348 (2012).

35 S.P. Muller, M.F. Kijewski, S.C. Moore, B.L. Holman, "Maximum-likelihood estimation: a mathematical model for quantitation in nuclear medicine," J Nucl Med **31**, 1693-1701 (1990).

36 M.K. Whitaker, E. Clarkson, H.H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods," Optics express **16**, 8150-8173 (2008).

37 S. Richard, E. Samei, "Quantitative breast tomosynthesis: from detectability to estimability," Medical Physics **37**, 6157-6165 (2010).

38 J.M. Wilson, O.I. Christianson, S. Richard, E. Samei, "A methodology for image quality evaluation of advanced CT systems," Medical Physics **40**, 031908 (2013).

39 K.L. Boedeker, V.N. Cooper, M.F. McNitt-Gray, "Application of the noise power spectrum in modern diagnostic MDCT: part I. Measurement of noise power spectra and noise equivalent quanta," Phys Med Biol **52**, 4027-4046 (2007).

40 J.B. Solomon, O. Christianson, E. Samei, "Quantitative comparison of noise texture across CT scanners from different manufacturers," Medical Physics **39**, 6048-6055 (2012).

41 J.M. Boone, "Determination of the presampled MTF in computed tomography," Medical Physics **28**, 356-360 (2001).

42 P.F. Judy, "The line spread function and modulation transfer function of a computed tomographic scanner," Medical Physics **3**, 233-236 (1976).

43 R.T. Droege, R.L. Morin, "A practical method to measure the MTF of CT scanners," Medical Physics **9**, 758-760 (1982).

44 E. Samei, O. Christianson, J. Chen, Z. Yang, G. Saiprasad, A. Dima, J. Filliben, A. Peskin, E. Siegel, "Task-Based Image Quality of CT Iterative Reconstruction Across Three Commercial Implementations," in *Medical Physics*, *Vol. 39* (AAPM, Charlotte, NC, 2012), pp. 4016.

45 C.C. Brunner, I.S. Kyprianou, "Material-specific transfer function model and SNR in CT," Phys Med Biol **58**, 7447-7461 (2013).

46 O. Christianson, J.J.S. Chen, Z. Yang, G. Saiprasad, A. Dima, J.J. Filliben, A. Peskin, C. Trimble, E.L. Siegel, E. Samei, "An Improved Index of Image Quality for Task-Based Performance of CT Iterative Reconstruction Across Three Commercial Implementations," Radiology, Accepted (2013).

47 E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M.J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D.M. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, S.J. Shepard, R.A. Uzenoff, J. Wang, C.E. Willis, "Assessment of Display Performance for Medical Imaging Systems," Report of the American Association of Physicists in Medicine (AAPM) Task Group 182005).

48 M.C. McDonald, H.K. Kim, J.H. Henry, I.A. Cunningham, "A novel method to measure the zero-frequency DQE of a non-linear imaging system," in *SPIE Medical Imaging*, *Vol. 7961* (SPIE, Orlando, FL, 2011), pp. 79610C.

49 E. Samei, M.J. Flynn, D.A. Reimann, "A method for measuring the presampled MTF of digital radiographic systems using an edge test device," Medical Physics **25**, 102-113 (1998).

50 A. Bharath, M. Petrou, *Next generation artificial vision systems : reverse engineering the human visual system*. (Artech House, Boston, 2008).

51 A.D.A. Maidment, M. Albert, "Conditioning data for calculation of the modulation transfer function," Medical Physics **30**, 248 (2003).

52 T. Flohr, K. Stierstorfer, R. Raupach, S. Ulzheimer, H. Bruder, "Performance evaluation of a 64-slice CT system with z-flying focal spot," Rofo **176**, 1803-1810 (2004).

53 M.M. Thornton, M.J. Flynn, "Measurement of the spatial resolution of a clinical volumetric computed tomography scanner using a sphere phantom," in *SPIE Medical Imaging*, *Vol. 6142* (SPIE, San Diego, CA, 2006), pp. 61421Z.

54 E. Samei, S. Murphy, S. Richard, "Assessment of multi-directional MTF for breast tomosynthesis," Physics in Medicine and Biology **58**, 1649-1661 (2013).

55 J. Baek, N.J. Pelc, "Use of sphere phantoms to measure the 3D MTF of FDK reconstructions," in *SPIE Medical Imaging*, *Vol. 7961* (SPIE, Orlando, FL, 2011), pp. 79610D.

56 S. Richard, E. Samei, "Quantitative imaging in breast tomosynthesis and CT: comparison of detection and estimation task performance," Medical Physics **37**, 2627-2637 (2010).

57 L. Yu, S. Leng, L. Chen, J.M. Kofler, R.E. Carter, C.H. McCollough, "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: Impact of radiation dose and reconstruction algorithms," Medical Physics **40**, 041908-041909 (2013).

58 J. Solomon, E. Samei, "Are uniform phantoms sufficient to characterize the performance of iterative reconstruction in CT?," in *SPIE Medical Imaging*, *Vol. 8668* (SPIE, Orlando, FL, 2013), pp. 86684M.

59 D.J. Brenner, E.J. Hall, "Computed tomography--an increasing source of radiation exposure," N Engl J Med **357**, 2277-2284 (2007).

60 B. Chen, S. Richard, E. Samei, "Relevance of MTF and NPS in quantitative CT: towards developing a predictable model of quantitative performance," in *SPIE Medical Imaging*, *Vol. 8313* (SPIE, San Diego, CA, 2012), pp. 83132O.

61 B. Chen, O. Christianson, J. Wilson, E. Samei, "Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods," Medical Physics, In review (2014).

62 A.E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," J Opt Soc Am A Opt Image Sci Vis **11**, 1237-1242 (1994).

63 E. Samei, M.J. Flynn, W.R. Eyler, "Detection of Subtle Lung Nodules: Relative Influence of Quantum and Anatomic Noise on Chest Radiographs," Radiology **213**, 727-734 (1999).

64 B. Chen, E. Samei, "Development of a phantom-based methodology for the assessment of quantification performance in CT," in *SPIE Medical Imaging*, *Vol. 8668* (SPIE, Orlando, FL, 2013), pp. 86681E.

65 E. Samei, M.J. Flynn, W.R. Eyler, "Simulation of subtle lung nodules in projection chest radiography," Radiology **202**, 117-124 (1997).

66 A.E. Burgess, X. Li, C.K. Abbey, "Nodule detection in two component noise: Toward patient structure," P Soc Photo-Opt Ins **3036**, 2-13 (1997).

67 X. Li, E. Samei, D.M. Delong, R.P. Jones, A.M. Gaca, C.L. Hollingsworth, C.M. Maxfield, C.W. Carrico, D.P. Frush, "Three-dimensional simulation of lung nodules for paediatric multidetector array CT," Br J Radiol **82**, 401-411 (2009).

68 A. Winklehner, C. Karlo, G. Puippe, B. Schmidt, T. Flohr, R. Goetti, T. Pfammatter, T. Frauenfelder, H. Alkadhi, "Raw data-based iterative reconstruction in body CTA: evaluation of radiation dose saving potential," Eur Radiol **21**, 2521-2526 (2011).

69 D. Marin, R.C. Nelson, S.T. Schindera, S. Richard, R.S. Youngblood, T.T. Yoshizumi, E. Samei, "Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm--initial clinical experience," Radiology **254**, 145-153 (2010).

70 P. Therasse, S.G. Arbuck, E.A. Eisenhauer, J. Wanders, R.S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A.T. van Oosterom, M.C. Christian, S.G. Gwyther, "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada," J Natl Cancer Inst **92**, 205-216 (2000).

71 World Health Organization., *WHO handbook for reporting results of cancer treatment*. (World Health Organization ; sold by WHO Publications Centre USA, Geneva; Albany, N.Y., 1979).

72 A.C. Silva, H.J. Lawder, A. Hara, J. Kujak, W. Pavlicek, "Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm," AJR Am J Roentgenol **194**, 191-199 (2010).

73 M. Das, J. Ley-Zaporozhan, H.A. Gietema, A. Czech, G. Muhlenbruch, A.H. Mahnken, M. Katoh, A. Bakai, M. Salganicoff, S. Diederich, M. Prokop, H.U. Kauczor, R.W. Gunther, J.E. Wildberger, "Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners," Eur Radiol **17**, 1979-1984 (2007).

74 J.P. Ko, H. Rusinek, E.L. Jacobs, J.S. Babb, M. Betke, G. McGuinness, D.P. Naidich, "Small pulmonary nodules: volume measurement at chest CT--phantom study," Radiology **228**, 864-870 (2003).

75 T.W. Way, H.P. Chan, M.M. Goodsitt, B. Sahiner, L.M. Hadjiiski, C. Zhou, A. Chughtai, "Effect of CT scanning parameters on volumetric measurements of pulmonary nodules by 3D active contour segmentation: a phantom study," Phys Med Biol **53**, 1295-1312 (2008).

76 H.X. Barnhart, D.P. Barboriak, "Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets," Transl Oncol **2**, 231-235 (2009).

77 C.S. Davis, *Statistical methods for the analysis of repeated measurements*. (Springer, New York, 2002).

78 K. Usuda, Y. Saito, M. Sagawa, M. Sato, K. Kanma, S. Takahashi, C. Endo, Y. Chen, A. Sakurada, S. Fujimura, "Tumor Doubling Time and Prognostic Assessment of Patients with Primary Lung-Cancer," Cancer **74**, 2239-2244 (1994).

79 W.J. Kostis, D.F. Yankelevitz, A.P. Reeves, S.C. Fluture, C.I. Henschke, "Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT," Radiology **231**, 446-452 (2004).

80 M. Petrou, L.E. Quint, B. Nan, L.H. Baker, "Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology," AJR Am J Roentgenol **188**, 306-312 (2007).

81 S. Sirohey, "LungVCAR-A technical description. Available via http://www.gehealthcare.com/usen/ct/clin_app/docs/9754WhtPprTech_4.pdf," (2005).

82 N. Petrick, H.J. Kim, D. Clunie, K. Borradaile, R. Ford, R. Zeng, M.A. Gavrielides, M.F. McNitt-Gray, Z.Q. Lu, C. Fenimore, B. Zhao, A.J. Buckler, "Comparison of 1D, 2D, and 3D Nodule Sizing Methods by Radiologists for Spherical and Complex Nodules on Thoracic CT Phantom Images," Acad Radiol **21**, 30-40 (2014).

Biography

Education

Duke University, Durham, NC 2014
Ph.D. in Medical Physics

Tsinghua University, Beijing, China 2008
B.E. in Engineering Physics, with Honors

National Tsing Hua University, Hsinchu City, Taiwan 2007
Exchange Student in Engineering and System Science

Honors and Awards

Carey E. Floyd Jr. Graduate Fellowship (*Top Research Honor*) , Duke Univ. 2013
Honorable Mention Poster Award, SPIE Medical Imaging Conference 2010
Carestream Health Fellowship, Duke Univ. 2009
Outstanding College Graduate of Beijing (*Top 5%*), Beijing Municipal Comm. of Edu. 2008
Outstanding Graduate of Tsinghua University, Tsinghua Univ. 2008
POSCO Asia Fellowships (*Top 5%*), Tsinghua Univ. 2005 - 2007
Outstanding Leader, Tsinghua Univ. Student Association 2006
Finalist (*Top 20*), 22nd Beijing College Student Physics Contest 2005
Huang Xianru Scholarship for Outstanding Academic Performance, Tsinghua Univ. 2005

Publications

Peer-reviewed articles

1. **Chen B**, Samei E. Developing a Prediction Model for Volume Quantification Performance in Computed Tomography. In preparation.
2. **Chen B**, Ramirez Giraldo JC, Samei E. Task-specific Evaluation of Iterative Reconstructions in CT (IRIS and SAFIRE). In preparation.
3. **Chen B**, Christianson O, Wilson J, et al. Assessment of Volumetric Noise and Resolution Performance for Linear and Nonlinear CT Reconstruction Methods. In review.
4. **Chen B**, Barnhart H, Richard S, et al. Volumetric Quantification of Lung Nodules in CT with Iterative Reconstruction (ASiR and MBIR). *Medical Physics*, 40(11), 111902 (2013). **(FEATURED ARTICLE)**
5. **Chen B**, Marin D, Richard S, et al. Precision of Iodine Quantification in Hepatic CT: Effects of Iterative Reconstruction with Various Imaging Parameters. *Am J Roentgenol*.

2013; 200(5):W475-W82.

6. Samei E, Li X, **Chen B**, Reiman R. The Effect of Dose Heterogeneity on Radiation Risk in Medical Imaging. *Radiat Prot Dosimetry*. 2013; 155(1):42-58.
7. **Chen B**, Barnhart H, Richard S, et al. Quantitative CT: Technique Dependence of Volume Estimation on Pulmonary Nodules. *Phys Med Biol*. 2012;57(5):1335-48.
8. Husarik DB, Marin D, Samei E, Richard S, **Chen B**, et al. Radiation Dose Reduction in Abdominal Computed Tomography During the Late Hepatic Arterial Phase Using a Model-Based Iterative Reconstruction Algorithm: How Low Can We Go? *Investigative Radiology*. 2012;47(8):468-74.
9. **Chen B**, Shorey J, Saunders RS, et al. An Anthropomorphic Breast Model for Breast Imaging Simulation and Optimization. *Academic Radiology*. 2011;18(5):536-46. **(ISSUE HIGHLIGHT)**

Full length conference proceedings

1. **Chen B**, Samei E. Development of a phantom-based methodology for the assessment of quantification performance in CT. *Proc. SPIE* 8668, Medical Imaging 2013: Physics of Medical Imaging, 86681E.
2. **Chen B**, Richard S, Samei E. Relevance of MTF and NPS in quantitative CT: towards developing a predictable model of quantitative performance. *Proc. SPIE* 8313, Medical Imaging 2012: Physics of Medical Imaging, 83132O.
3. **Chen B**, Richard S, Christianson O, et al. CT performance as a variable function of resolution, noise, and task property for iterative reconstructions. *Proc. SPIE* 8313, Medical Imaging 2012: Physics of Medical Imaging, 83131K
4. Kiarashi N, Yuan Lin Y, Segars WP, Ghate SV, Ikejimba L, **Chen B**, et al. Development of a dynamic 4D anthropomorphic breast phantom for contrast-based breast imaging. *Proc. SPIE* 8313, Medical Imaging 2012: Physics of Medical Imaging, 83130C
5. Ikejimba L, Kiarashi N, Lin Y, **Chen B**, et al. Task-based strategy for optimized contrast enhanced breast imaging: analysis of six imaging techniques for mammography and tomosynthesis. *Proc. SPIE* 8313, Medical Imaging 2012: Physics of Medical Imaging, 831309
6. **Chen B**, Samei E, Colsher J, et al. Precision of iodine quantification in hepatic CT: effects of reconstruction (FBP and MBIR) and imaging parameters. *Proc. SPIE* 7961, Medical Imaging 2011: Physics of Medical Imaging, 79610L.
7. **Chen B**, Marin D, Samei E. A new iodinated liver phantom for the quantitative evaluation of advanced CT acquisition and reconstruction techniques. *Proc. SPIE* 7961, Medical Imaging 2011: Physics of Medical Imaging, 796161.
8. **Chen B**, Richard S, Barnhart H, Colsher J, Amurao M, Samei E. Quantitative CT:

technique dependency of volume assessment for pulmonary nodules. *Proc. SPIE* 7622, Medical Imaging 2010: Physics of Medical Imaging, 76222W. **(HONORABLE MENTION POSTER AWARD)**

9. Richard S, **Chen B**, Samei E. Extending the detectability index to quantitative imaging performance: applications in tomosynthesis and CT. *Proc. SPIE* 7622, Medical Imaging 2010: Physics of Medical Imaging, 76220Z
10. Samei E, Li X, **Chen B**, et al. The myth of mean dose as a surrogate for radiation risk? *Proc. SPIE* 7622, Medical Imaging 2010: Physics of Medical Imaging, 76220T
11. Thompson J, **Chen B**, Richard S, et al., "Wide-angle breast tomosynthesis: initial comparative evaluation. *Proc. SPIE* 7622, Medical Imaging 2010: Physics of Medical Imaging, 76220C

Presentations

1. **Chen B**, Samei E. Development of a phantom-based methodology for the assessment of quantification performance in CT. SPIE Medical Imaging. Orlando, FL 2013.
2. **Chen B**, Richard S, Samei E. Relevance of MTF and NPS in quantitative CT: towards developing a predictable model of quantitative performance. SPIE Medical Imaging. San Diego, CA 2012.
3. **Chen B**, Richard S, Christianson O, et al. CT performance as a variable function of resolution, noise, and task property for iterative reconstructions. SPIE Medical Imaging. San Diego, CA 2012.
4. **Chen B**, Richard S, Zhou X, Samei E. Iodine Quantification in Liver CT: the Impact of Iterative Reconstruction (IRIS and SAFIRE), Dose, Slice Thickness, and Patient Size. RSNA. Chicago, IL 2011.
5. Richard S, **Chen B** (presenter), Zhou X, Samei E. Performance Evaluation and Comparison of Filtered Backprojection with Clinical and Investigational Iterative Algorithms (IRIS and SAFIRE). RSNA. Chicago, IL 2011.
6. Richard S, **Chen B** (presenter), Samei E. A Novel Method for Predicting the Performance of CT Lung Nodule Volume Quantification. RSNA. Chicago, IL 2011.
7. **Chen B**, Samei E, Colsher J, et al. Precision of iodine quantification in hepatic CT: effects of reconstruction (FBP and MBIR) and imaging parameters. SPIE Medical Imaging. Orlando, FL 2011.
8. **Chen B**, Marin D, Samei E. A new iodinated liver phantom for the quantitative evaluation of advanced CT acquisition and reconstruction techniques. SPIE Medical Imaging. Orlando, FL 2011.
9. **Chen B**, Samei E, Richard S, Husarik D, Marin D, Nelson R. Quantification of Iodine

Concentration in Abdominal CT: Effects of Reconstruction (FBP, ASiR, and MBiR) and Dose. RSNA. Chicago, IL 2010.

10. **Chen B**, Richard S, Barnhart H, Colsher JG, Maxwell A, Samei E. Volumetric Quantification of Lung Nodule in CT: Effects of Reconstruction Algorithm (FBP, ASiR, and MBiR), Dose and Slice Thickness. RSNA. Chicago, IL 2010.
11. **Chen B**, Richard S, Barnhart H, Colsher J, Amurao M, Samei E. Quantitative CT: technique dependency of volume assessment for pulmonary nodules. SPIE Medical Imaging. San Diego, CA 2010.